A MATLAB TOOL TO QUANTIFY COLONY SIZE DISTRIBUTIONS IN VITRO AND IN VIVO

by

Soumitra Bhoyar

A dissertation submitted to The Johns Hopkins University

in conformity with the requirements for the degree of

Master of Science in Engineering

Baltimore, Maryland

May, 2018

Abstract

Ninety percent of cancer-related deaths are due to metastasis, hence improved methods to understand and model metastasis are required. Current methods of analyzing metastatic events in experimental animal models do not provide information on colony size distributions, and suffer from the inability to segment micrometastases. Obtaining quantitative metrics in vivo would be particularly useful in settings involving fluorescent cells, which are becoming increasingly widespread for in vitro and in vivo applications, and are important tools in identifying roles of specific proteins or genes in the metastatic cascade. Quantification of metastatic colony size distributions would also find applications in investigating clonal competition in genetically heterogenous tumors and visualizing genetic exchange between cancer cells. Furthermore, there are a limited number of mathematical models that describe metastasis (as opposed to several that focus on tumor growth).

We have developed an image-processing based method, designed using MATLAB that effectively obtains quantitative data from fluorescent cell colonies both in vitro and in vivo. The method is sensitive enough to segment lung micrometastases consisting of only 4-5 cells, which makes it suitable for mathematical modelling of cancer metastasis. The lower detection limit

(compared to bioluminescence imaging or computational tomography, which are often used to validate mathematical models), higher resolution and speed of our method would assist in obtaining quantitative data for the purpose of modelling.

Our methodology can also be applied to in vitro systems as well. While clonogenic assays are often used to obtain growth potential, the assay described in this work provides confluence, colony size distributions, colony-forming potential and synergistic/ co-culture effects in fluorescently labelled cells. Also, the method provides distributions of immunostained proteins of interest between heterogenous cell populations, for instance the proliferation marker ki67. The method is fast and accurate even for densely spaced colonies, and we plan to release the MATLAB scripts as freeware for general use.

Thesis Committee

Primary Readers

Daniele Gilkes (Primary Advisor)
Assistant Professor
Department of Chemical and Biomolecular Engineering
Johns Hopkins University

Denis Wirtz
Theophilus Halley Smoot Professor
Department of Chemical and Biomolecular Engineering
Johns Hopkins University

Acknowledgments

I would like to thank my research advisor, Dr. Daniele Gilkes, for her insight and guidance throughout my MSE program at Johns Hopkins. I am grateful to her for constantly helping me develop new skills and grow as a researcher. I would also like to acknowledge all members of the Gilkes Lab for creating a supportive and enjoyable atmosphere to work in. Thank you Inês Godet for your contribution to this thesis, and for valuable discussions throughout the way.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Cancer Metastasis

Cancer metastasis is the term used to describe the dissemination of cancer cells to distant organs from the primary tumor. Ninety percent of cancer-related deaths are due to metastasis rather than due to the primary tumor, a proportion that has not changed significantly over the last five decades [1]. Metastasis requires a series of interrelated events to take place, which are described as the metastatic cascade. The following steps are involved: cancer cell detachment from the primary tumor mass and invasion of the surrounding tissue, intravasation into local blood or lymphatic vessels (i.e. the circulatory system), extravasation through capillary beds at a secondary site, invasion of the secondary organ and subsequent proliferation. Furthermore, tumor cells may induce angiogenesis at the secondary site, form secondary metastases (metastases of metastases), and evade the host immune system throughout the cascade [2] [1].

1.2 Experimental Analysis of Metastases

1.2.1 Mice Models

Experimental mouse models are a popular method of studying metastasis. In order to study the metastatic process in such experimental models, the ability to quantify the extent of the metastases is needed. However, current methods of quantitatively analyzing metastases *in vivo* do not provide sufficient quantitative information pertaining to metastatic colonies or their spatial distributions at the organ-level [3].

1.2.2 Hematoxylin & Eosin Staining

Traditionally, Hematoxylin & Eosin (H&E) stained tissue sections have been used to visualize metastatic regions in organs. To obtain quantification of metastases in lungs, the following simplified procedure is used. After sacrificing the mouse, most of the rib cage is removed and the lungs are inflated by gently injecting with 4% Paraformaldehyde (PFA). Next, the lungs and heart are removed, submerged in PFA and embedded in paraffin after which H&E staining is carried out. After the paraffin-embedded tissue sections are mounted on slides, they may be scanned such that the whole tissue section is captured digitally. Software programs such as ImageJ, NIS-Elements and CellProfiler exist that calculate overall tissue area based on freehand selection (which provides us a measure the total tissue area). Finally, areas containing metastatic cells are identified and traced manually. If fluorescent markers are being used, the fluorescent regions may be identified by binarization,

i.e. converting pixel intensities into binary true/false, gated according to a threshold. The threshold can be set manually or an optimum threshold may be obtained by an algorithm, such that true fluorescent pixels are accepted and pixels containing autofluorescence or background signal are not.

From the above exercise it is possible to get total metastatic area, number of metastases, total tissue area, and hence average area of each metastasis as well as the density of metastases in lung tissue. Region-Of-Interest (ROI) tracing on sections can provide an estimate of individual metastatic colony sizes . However this is highly labor-intensive and time consuming, especially when dealing with high metastatic loads and a large number of sample images. Furthermore, the user has to visually judge what regions constitute a 'metastasis', taking care to avoid regions of immune infiltration and regions around blood vessels or bronchioles, which exhibit higher packing density and might appear similar to tumor cells [4].

1.2.3 Quantitative Polymerase Chain Reaction

Quantitative polymerase chain reaction (qPCR) is another tool that is used to quantify overall metastatic burden in organs, and has been applied to both xenograft and syngeneic experimental models. In xenograft models, where human tumor cells are transplanted onto a host belonging to a different species, detection and quantization of human *alu* sequences from host organ DNA extracts is the basis for quantifying metastatic load [5–7]. In syngeneic models this method cannot be used since the tumor and host cells have the same genome. For syngeneic models, a recent study introduced a modified

luciferase gene (*Luc2*) into tumor cells using a lentivirus. By using a primer designed for amplifying a luciferase gene fragment, it was possible to quantify Luc2-positive tumor cells, with a sensitivity of one tumor cell per 10,000 [3]. Another study showed that it is possible in certain cell lines to quantify metastatic load in syngeneic models using qPCR based on differences in expression of signature genes between metastatic cells and normal cells [8].

While qPCR based methods are sensitive and can measure metastatic loads in a host organ, they require homogenized tissues, or tissue extracts, and hence no information on metastatic colony sizes is obtained.

1.2.4 Bioluminescence Imaging

Bioluminescence imaging (BLI) is a popular method of quantifying tumor and metastatic loads based on measuring fluorescence values. Cancer cells are genetically modified to express the enzyme luciferase before being introduced into the animal. Subsequently, at a suitable time point, the substrate, luciferin, is injected in the animal. Upon coming into contact with the substrate, luciferase-expressing cancer cells emit photons through the luciferase protein reaction. 'Luciferin' and 'Luciferase' are general terms for the substrate and enzyme which react to produce light emission in bioluminescent organisms. For research applications the most commonly used enzyme/substrate pair is the one present in the male firefly *Photinus pyralis*. Mice are injected with luciferin prior to imaging, and are imaged using a Xenogen imager.

While it requires an expensive substrate and specialized imaging equipment, an advantage of BLI is that it is a dynamic assay and can be conducted

over time, with mice anaesthetized before every measurement, making it suitable for visualizing tumor progression and metastasis. While BLI is suitable for qualitative or yes/no applications, it does not allow true quantification of the target signal. This is because the signal depends on a variety of factors aside from luciferase expression, such as depth, and local concentrations of luciferin, ATP and oxygen. Furthermore, any signal received is significantly weighted towards the surface due to the rapid signal attenuation that occurs as the photons pass through tissue [9, 10].

1.3 Mathematical Modelling of Cancer

1.3.1 Overview

Mathematical modelling of cancer has been on the rise in past years and has proved to be an important tool in understanding the disease. Mathematical models have been employed to shed light on different processes in tumor growth, suggest treatment options, devise experiments and re-evaluate prognoses [11]. Mathematical models are useful for integrating disparate experimental data, in order to understand the overall process of cancer progression. Since experiments are usually designed with a narrow focus, limited to one or few steps of tumor growth and the metastatic cascade, it falls on mathematical models to provide higher level insights of these processes. However, as of today, modelling has been mostly limited to primary tumor-growth, as opposed to metastasis [12].

Several models have been proposed to predict tumor growth and dynamics, from the earliest ones that were used as predictors of tumor size, to those

that attempt to model the various effects of signal transduction pathways. Of the former kind, the phenomenological Gompertz model has remained most popular despite its empirical nature, due to its accurate prediction of tumor size and ease of application to different data sets [13]. Recent approaches involve developing models that account for the effects of various factors involved in tumor cell growth, such as the microenvironment, vasculature and angiogenesis, the impact of therapeutics, and cancer stem-cells [14–18].

1.3.2 Modelling Cancer Metastasis

Modelling metastatic processes is a challenge due to the complexity of the metastatic cascade, hence while several models exist that predict tumor growth or model interactions that occur in the tumor microenvironment, fewer have been developed, especially recently, that predict metastatic spreading [11, 19]. One of the motivations for a predictive model for metastasis is to obtain metastatic colony size distributions. Specifically, to estimate the formation of micrometastases that are only a few cells large, and which are below the detection limit of clinical diagnostics.

One study developed a deterministic, dynamical model to obtain the size distribution of metastases, and validated it using colony size measurements obtained by computed tomography (CT) scans of a patient with hepatocellular carcinoma before and after therapy. It was discussed that quantitative prediction of metastatic colony size distribution is important in the clinical setting since colony sizes and vascularity, both of which are estimated by the model, play a role in determining the effects of chemotherapy [20]. The same model

was fitted on data obtained in a follow-up study that examined 3,500 patients for metastatic relapse, and it was found to accurately predict metastatic risk. This follow-up study, while demonstrating the reliability of the former model, provided no information about the growth dynamics of the metastases [21]. .

Similar to the deterministic model cited above, some stochastic models have been developed that predict the state of micrometastases present in the patient, by modelling the colony size distributions in the patient at a given time. [22–24]. Such patient-specific data is unavailable to clinicians at this time, and represents a step toward having personalized information of the disease [12]. While powerful, these studies are based on a single patient and involves cancer progression in a single organ, hence a more rigorous validation would seem to be the next logical step.

A more recent study attempted to perform such a rigorous validation using an orthotopic xenograft in mice. It is logical to use animal experiments for model validation since obtaining the equivalent data clinically would be difficult due to ethical reasons and the long time scales involved in cancer progression. In this study, quantification of metastatic spreading over time was done using in vivo 3D Bioluminescence Imaging (BLI). Due to the limitations of BLI discussed in the previous section, the experimenters could measure only the tumor size and total metastatic load, with no quantification of the individual metastatic foci or their impacts on metastasis. Another concern is that the BLI measurements were calibrated by obtaining an *in vitro* cell-to-signal ratio, which represents a source of experimental error, since signal attenuation through mouse tissue and variation of the luciferase reaction

between different organs of the host (as discussed previously) may affect the experimental (*in vivo*) cell-to-signal ratio [19].

1.4 Fluorescent Cells And Cancer Research

Fluorescent proteins, such as the Green Fluorescent Protein (GFP) and other related proteins, do not require substrates or cofactors to emit light due to the presence of internal chromophores. Cells expressing or tagged with fluorescent proteins have found many different applications in cancer research over the past two decades [25].

In one study fluorescent proteins were used to visualize *in vivo* genetic exchange between prostate cancer cells, and it was observed that this led to enhanced metastatic capability. Two parental cell lines were used in this study, separately expressing GFP and RFP. Cells that had undergone genetic exchange expressed both proteins and appeared yellow [26]. The same lab used this principle to investigate metastatic clonality in other cancer cell lines [27]. A similar study was performed with a high metastatic and a low metastatic human osteosarcoma cell line showed an increase in metastatic potential in the low metastatic cell line when co-transplanted with the high metastatic cell line, but not when transplanted separately. The system was color-coded similar to the one described above, which enabled the experimenters to determine the origin of metastatic colonies in the host organs [28]. Aside from clonality, a dual-colour fluorescence based imaging system has also been used to investigate stem-cell like behaviour in cancer cell lines [29].

Fluorescent proteins have been used in several ways to visualize cells and

their behaviour in mice via the creation of transgenic mice. Furthermore, transgenic reporter mice have been developed which are able to label specific cellular characteristics, such as gene expression or cell cycle progression [30]. For example, a reporter mouse expressing GFP-tagged β-actin, controlled by a keratin-14 (K14) promoter was generated [31]. This allowed visualization of K14+ tumor cells that coexpressed an mTomato reporter construct. [32]. The tool described in this work can be used to investigate and quantify metastatic potential, growth and colony formation of the K14+ cells versus K14- cells both *in vivo* and *in vitro*. In an interesting study, a Cre-based fluorescent reporter system was developed that changed color upon epithelial-to-mesenchymal transition [33]. Cells which had undergone EMT were found to be more likely to lead to early metastatic events, and were more resistant to chemotherapy. In this study, overall metastatic loads and the fractional contributions by the dual colored cells were reported. The assay described in this work could be applied to this system to obtain colony level data *in vivo* and *in vitro*.

1.5 In Vitro Clonogenic Assay

The clonogenic assay was first described in 1956 in an experiment that analyzed the effect of x-rays on the colony-formation ability (clonogenicity) of a human cervical carcinoma cell line (HeLa cells). In the clonogenic assay, cells are fixed and stained by crystal violet, and counted manually [34]. The x-ray dose/ survival curves obtained from this assay (obtained by counting the number of viable colonies) represented a quantification of the cytotoxic effects of ionizing radiation. Since then the assay has been modified and applied to a

number of experiments and is currently a widely-used experiment for cancer biologists [35]. Some computational tools have been developed for the assay, for statistical analyses [36] and automated colony segmentation [37]. While the clonogenic assay can quantify differences in sensitivity to cytotoxic agents, it is not suitable for co-culturing different cell lines and cannot visualize proteins of interest. Also, since the assay involves crystal-violet staining, automated segmentation algorithms are based on colony area as opposed to cell count (although the discrepancy between the two counts is mitigated by taking into account the intensity of the stain).

1.6 Digital Pathology and Image Processing

Digital pathology has gained increasing importance over the past decade with increases in computational power and the advent of sophisticated image analysis techniques. While traditional visual microscopy based pathology is the current gold-standard, it can be prone to lack of standardization or variation between pathologists [38, 39], and is time consuming [40]. These are the principal issues which digital pathology seeks to overcome. While digital pathology is slowly gaining relevance in the clinical setting, there are some technical issues which remain to be tackled [41–43]. For example, three-dimensional cell clusters or dense smears make digital pathology difficult [44]. In research applications involving fluorescently labelled cells, some of the limitations encountered in the clinical setting do not apply. Fluorescently labelled cells are a popular choice *in vitro* and are becoming increasingly widespread in cancer research for *in vivo* experiments (as described in the

previous section). From a digital pathology point-of-view, these cells have an advantage in that they are easier to identify and can be segmented into colonies with greater ease than clinical specimens stained with colorimetric dyes such as H&E. This facilitates the application of image analysis techniques in obtaining quantitative data from histological sections.

1.7 Motivation

The methodology described in this work to analyze experimental metastases provides sharper, more granular data than existing techniques. So far, there has been no effort in quantifying metastases at the colony level, partly due to the lack of sensitive enough methods. However, images obtained through fluorescence microscopy are suitable for image analysis techniques to quantify information directly from fluorescence micrographs with speed and accuracy.

There is a need for experimental metastatic data for creating and validating effective mathematical models of cancer metastasis, of which there are currently few available. Presently, experimental data for mathematical models is obtained from BLI or similar methods, which provide no data at the colony level. Furthermore, this methodology would be useful for quantification and modelling of synergistic/ non-autonomous effects, and investigating clonal competition in cancer cell lines, both *in vitro* and *in vivo*. The *in vitro* assay has the advantage of being fast and experimentally simple, while giving quantitative information on proliferation, dynamics of colony formation and protein expression over time, while identifying co-culture effects, in a single experiment.

Finally, with the popularity of transgenic reporter mice and fluorescent cells reporting expression of proteins- or genes-of-interest, we require a fast and simple way of quantifying differences in proliferation, colony formation and metastatic potential of different cell lines.

In our lab, the *in vivo* methodology was used to obtain differences in metastatic colony size distributions of MDA-MB-231 tumor cells. The tumor cells used contained a Hypoxia Inducible Factor (HIF) - regulated reporter plasmid construct, a sequence that causes the cell to permanently switch from expressing DsRed to GFP upon exposure of the cell to hypoxia in the primary tumor (unpublished work). Hence, upon the onset of tumor hypoxia, a portion of the tumor cells shift from expressing DsRed to GFP, which in turn leads to metastatic colonies in mice organs which are either DsRed+ or GFP+. Using our tool, we identified and quantified differences in size distributions between metastases formed by DsRed+ cells and those formed by GFP+ cells.

Chapter 2

Materials and Methods

2.1 Cell Culture

MCF7 and MDA-MB-231 cells were cultured in Dulbecco's Minimum Essential Media (DMEM) (Sigma-Aldrich, St. Louis MI - USA) supplemented with 10% Fetal Bovine Serum (FBS) and 1% penicillin-streptomycin (Invitrogen, Carlsbad) in a humidified atmosphere of 5% CO_2 at 37°C.

2.2 Fluorescent Cell Lines

MDA-MB-231 and MCF-7 cell lines DsRed- and GFP- expressing were generated by transfecting them with two independent lentiviral vectors. Vector 1 expresses a DsRed reporter with a stop codon flanked by tandem loxP sites ('floxed') and located in front an GFP gene. In the absence of Cre-recombinase, cells express a red fluorescent protein (DsRed) as shown in figure 2.1A. DsRed-expressing cell lines were transfected exclusively with vector 1. Deletion of DsRed by Cre recombinase results in the rapid loss of DsRed and the permanent activation of GFP expression. Vector 2 expresses the Cre gene under the

control of the constitutively activated U6 promoter and CMV enhancer (figure 2.1B). GFP-expressing cell lines were generated from the existent DsRed-expressing cells after transfection with vector 2. All cell lines were sorted using a SH800 Cell Sorter in sterile conditions, to guarantee purity and high-fluorescence.

To generate a hypoxia-reporter system, MDA-MB-231 cells already transfected with vector 1, were then transfected with vector 3. Vector 3 expresses the Cre gene under the control of a synthetic hypoxia reporter which was designed based on the putative HIF-binding sequence. Under hypoxic conditions, HIFs will drive the expression of Cre to excise the DsRed coding sequence in vector 1 (2.1C and D).

2.3 *In Vitro* Colony Formation Assay

2.3.1 Plating

DsRed-expressing cells (DsRed+) and GFP-expressing cells (GFP+) were used in our experiments. To model cell lines with different proliferation rates and clonogenicity, GFP+ cells were cultured in media containing 2 nM Paclitaxel for 48 hours prior to plating for the *in vitro* experiment, whereas DsRed+ cells were cultured in normal media. Following this, 2000 each of DsRed+ and Paclitaxel-precultured GFP+ cells were added to three wells of a 6 well-plate, for a total of 4000 cells per well, and grown in normal media with standard tissue culture. To the remaining three wells of the well-plate, 2000 each of DsRed+ and untreated GFP+ cells were added, to function as a control.

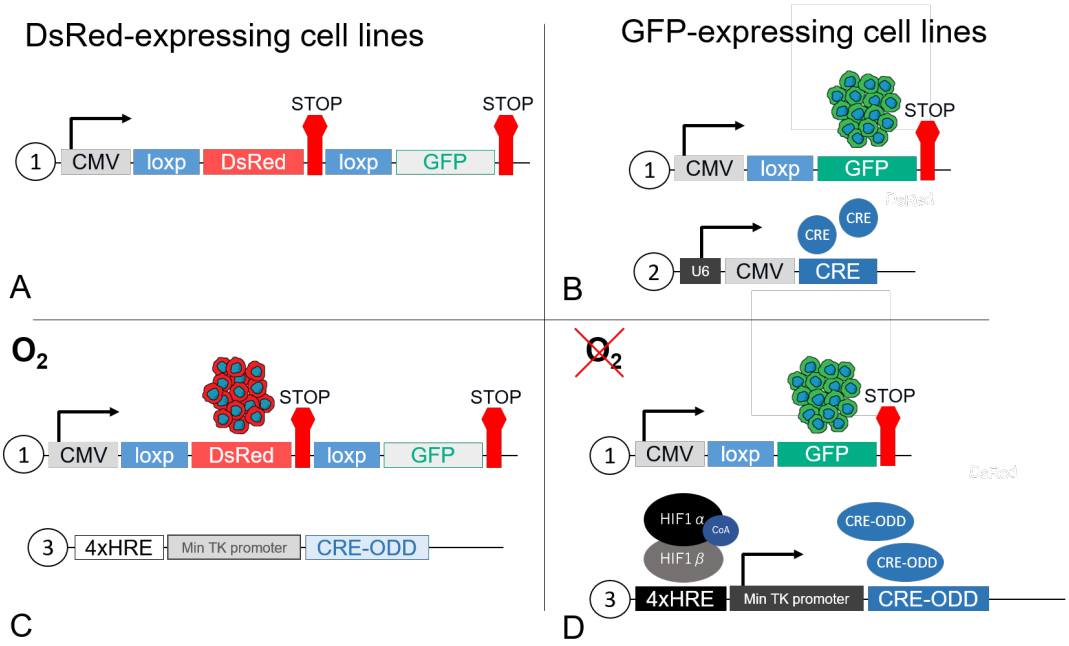

Figure 2.1: **A:**Vector 1 expressing floxed DsRed. This system is present in DsRed+ cells. **B:** Vectors 1 and 2. The action of vector 2 results in DsRed cleavage from vector 1. This system is present in GFP+ cells. **C:** Vectors 1 and 3 under normoxia. **D:** Vectors 1 and 3 under hypoxia, resulting in DsRed cleavage.

2.3.2 Live-cell vs. Fixed-cell Modes

We designed two modes of operation for the *in vitro* colony formation assay - live-cell and fixed-cell. The fixed-cell mode is an end-point assay that involves fixing with paraformaldehyde and allows for immunostaining of proteins of interest. The live-cell mode allows the user to observe and quantify colony growth in a given plate continuously over time. A full comparison is provided in the subsequent sections.

2.3.3 Fixing and Staining

This section applies to the fixed-cell mode only. A 6 well-plate was fixed at each time point and stained for Ki-67 as follows.

1. Culture media was removed
2. Cells were fixed by adding 4% paraformaldehyde (PFA) in phosphate-buffered-saline (PBS) for 15 minutes.
3. Triton-X (1%) was added for 5 minutes.
4. Bovine serum albumin (2% in PBS) was added for 30 minutes.
5. Immunostaining was performed with anti Ki-67 monoclonal antibody (Alexa Fluor 647 Mouse anti-Ki-67, BD Biosciences) for 90 minutes.
6. Nuclear staining was performed with a 5μg/mL Hoechst 33342 solution in PBS, for 15 minutes.

All wells were washed in PBS after each step. Fixed and stained cells were stored at 4°C under PBS.

2.3.4 Imaging

Plates were imaged using a Cytation 5 imaging reader (BioTek, Winooski, VT, USA) equipped with an incubated stage. For the live-cell mode, the stage was maintained at 37°C, and images of the same plate, at the same position were obtained in the RFP and GFP channels at timepoints of 1, 2 and 4 days. For the fixed-cell mode, separate plates prepared on day 0 were imaged under RFP, GFP, DAPI and CY5 channels on days 2, 4 and 5.

Plates were imaged at a magnification of 4x. A montage of images was used to cover approximately 80% of the well area, which provided a sufficient number of colonies to obtain size distributions. The images were stitched with NIS Elements software(Nikon Instruments Inc., Melville, NY, USA) and their dimensions reduced to 50% in order reduce processing time.

2.4 Image Processing and Analysis - *In Vitro*

2.4.1 Introduction

In this section we discuss the image-processing filters and algorithms used in obtaining quantitative data from the *in vitro* images generated in the previous section. All scripts were created in MATLAB 2017b (MathWorks Inc., Natick, MA, USA), using the Image Processing Toolbox available for MATLAB. The image processing steps of the fixed cell mode is summarized in figure 2.3, and described in detail in the subsections below (from 2.4.2 to 2.4.6). The algorithmically simpler live-cell mode is discussed in subsection 2.4.7. A flowchart of the steps involved is provided in figure

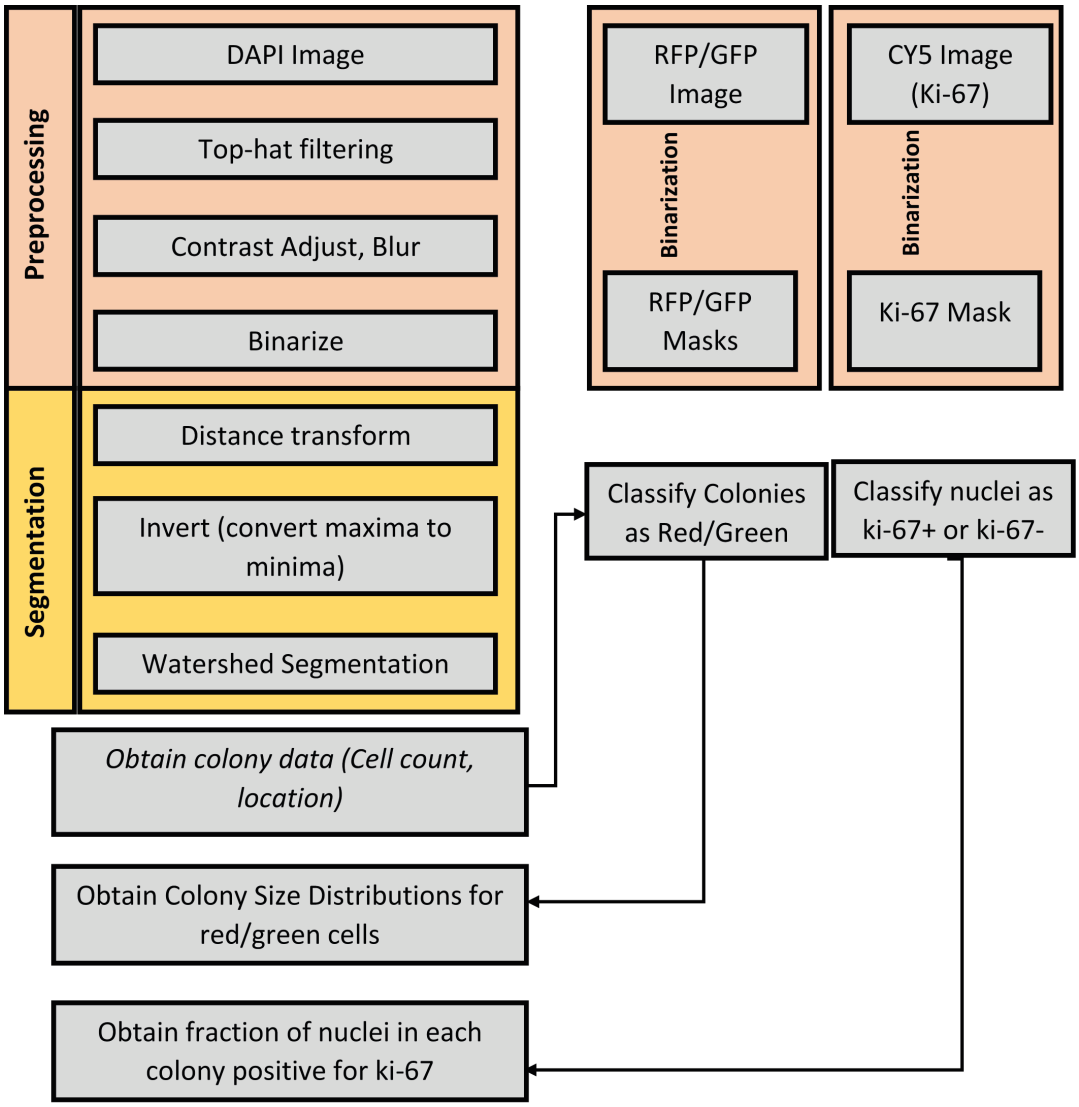

Figure 2.2: Flowchart of the image processing scheme employed in the *in vitro* experiment.

The script loads the RFP, GFP, DAPI and CY5 images generated in the imaging step in the fixed-cell mode and performs the operations described below.

2.4.2 Preprocessing

Figure 2.3 A: On the DAPI (and as well as the RFP and GFP) images, as a background correction step, top-hat filtering is used. This reduces background fluorescence, signal due to bubbles and lens artifacts. Gaussian filtering is used to eliminate jagged boundaries an obtain smooth colony-regions. A final contrast adjust step ensures that the fluorescent regions are above the threshold of segmentation. The results of the preprocessing steps are shown in Figure 2.4. After this step, the image is binarized.The binarized DAPI image identifies pixels that are occupied by cell colonies (as opposed to background).

2.4.3 Distance Transform

Figure 2.3 B: In the binarized image obtained in the previous step, a given pixel has the value '1' if it lies within a colony and '0' if not. A distance transform is then performed on the complement of the binary image, which effectively replaces the value of each pixel within a colony with the Euclidean distance to the nearest boundary. This image was inverted (dark regions were made bright and vice versa), and the background was made bright. The results of this treatment are shown in Figure 2.5. Observe the dark regions at the centres of each colony, which function as catchment regions for the watershed segmentation step mentioned below.

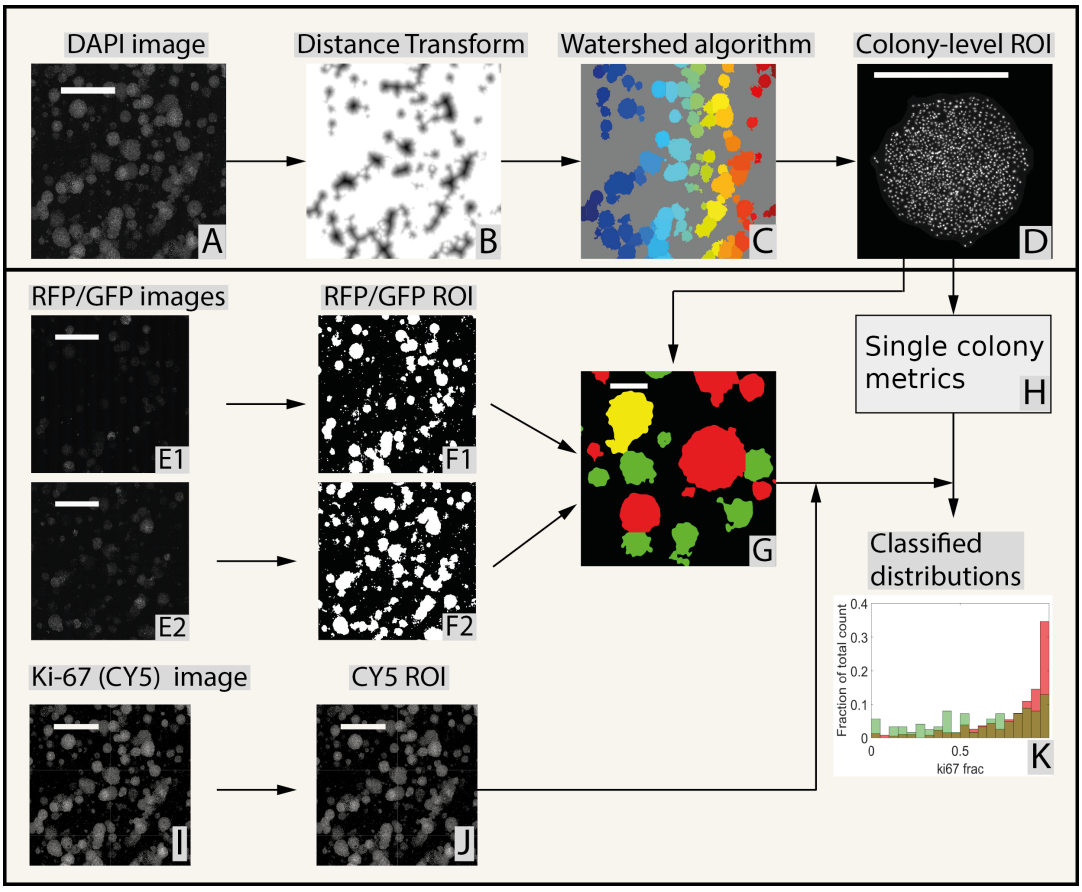

Figure 2.3: A: Nuclear staining (DAPI channel). **B:** Preprocessing steps and distance transform identifies watersheds. **C:** Segmentation of colonies by watershed algorithm. **D:** Individual colonies analyzed. **E:** DsRed and GFP signal captured in RFP and GFP channels respectively. **F:** DsRed+/GFP+ regions obtained by preprocessing steps and binarization. **G:** Segmented colonies classified as DsRed+/ GFP+ based on input from images F1 and F2. **G:** Cell count and area of individual colonies obtained from colony-level DAPI image. **I:** Ki-67 image in CY5 channel. **J:** Regions expressing Ki-67 obtained by binarization. **K:** Distribution of Ki-67 coverage, colony sizes obtained separately for populations of DsRed+ and GFP+ cells. **Scalebars:** 5 mm (A, E, I, J), 1 mm (G, I)

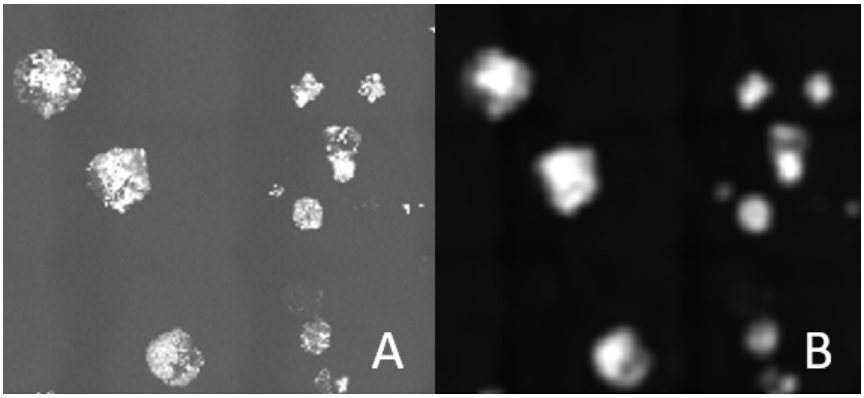

Figure 2.4: A: Raw stitched image obtained from plate reader. **B:** Image after preprocessing (top-hat filtering, gaussian filtering and contrast adjustment)

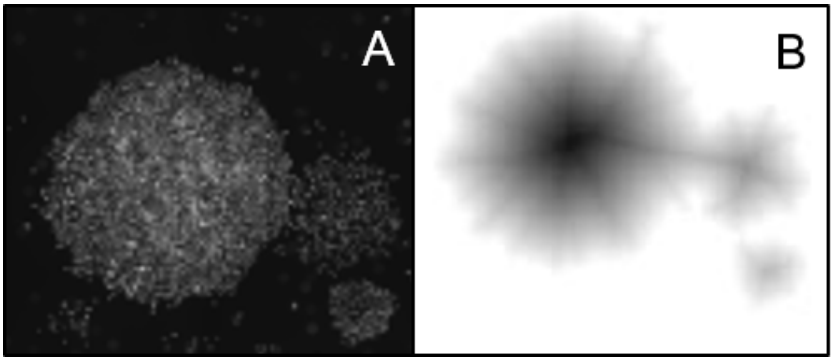

Figure 2.5: A: Raw stitched image showing one colony. **B:** Distance Transform applied and inverted.

This method takes advantage of the fact that in vitro colonies are roughly circular. We use an h-minima transform to suppress small minima which otherwise lead to oversegmentation.

2.4.4 Watershed Segmentation

Figure 2.3 C: The watershed transform is a popular method for image segmentation. The algorithm views the subject image as a topographical map, viewing bright pixels as 'high' and dark ones as 'low'. It segments the image

into catchment areas based on dark 'basins', which are then used as foci for segmenting objects corresponding to each basins [45, 46]. In our application, we use a distance transform to obtain 'high' or bright regions, invert them so that they become 'basins', and subsequently use them as foci for segmenting colonies *in vitro*. An example of the segmentation step is provided in figure 2.6, where a large group of connected colonies is accurately segmented into individual colonies. While the script works well with clustered or partially overlapping colonies, it cannot distinguish sheet-like areas or areas where multiple colonies have fused into a 'super-colony', which is also shown in the figure.

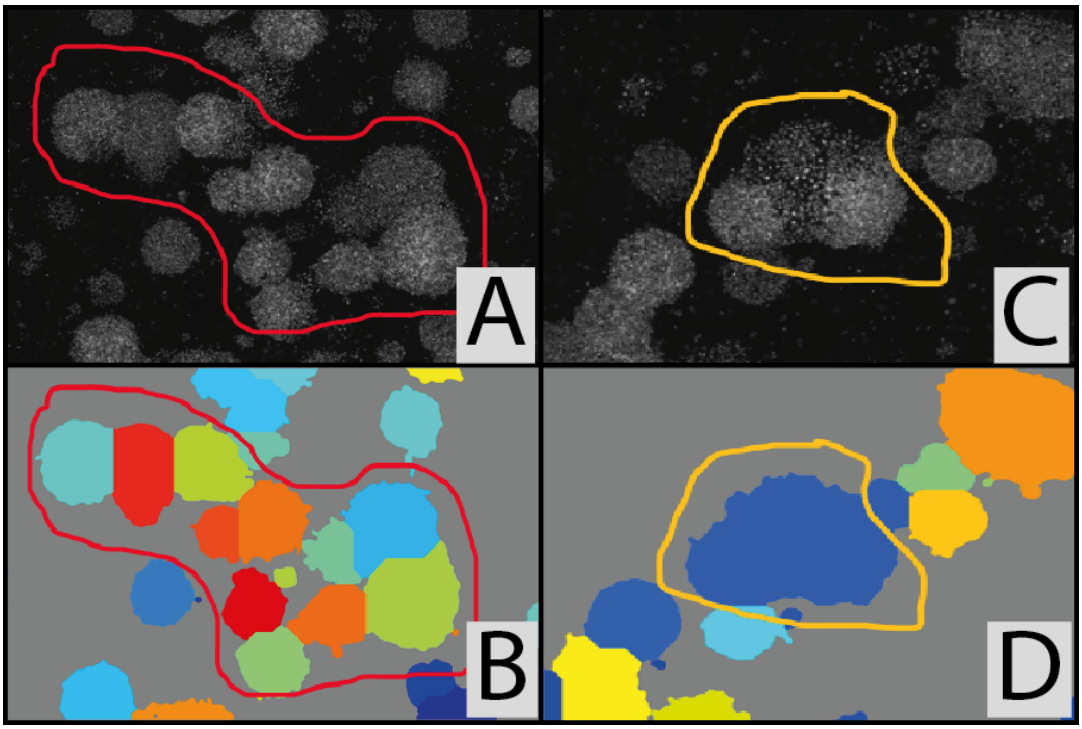

Figure 2.6: A: A group of connected colonies. **B:** Successful segmentation by watershed. **C:** A group of severely merged colonies. **D:** Failure to segment severely merged colonies

2.4.5 Colony-Level Analysis

Figure 2.3 D, H: The segmentation step provides us colony sizes. Furthermore, individual nuclei in with each colony are counted (algorithmically) based on the signal received from the nuclear-stained cells. The algorithm for segmenting nuclei closely resembles the colony segmentation step in that it also involves a distance transform followed by a watershed algorithm. Hence, the fixed cell assay provides colony cell counts as well as areas. The results of the segmentation can be seen in figure 2.7A-C.

Furthermore, based on input from the RFP/GFP channel images, segmented colonies may be classified as DsRED+ or GFP+, which is shown in figure 2.3 E-G. The signal from binarized GFP/RFP images determine whether a given colony (segmented in the previous steps) consists of pixels that are GFP+ or DsRed+, which is the basis for classifying colonies based on originating cell line. Hence, distributions of colony areas and cell counts can be obtained from each of the two populations (DsRed+ and GFP+), and quantitative differences between them may be analyzed.

2.4.6 Ki-67 Analysis

Figure 2.3 I-K: The fixed-cell assay enables the experimenter to obtain distributions of immunostained proteins of interest. In our assay we use the protein Ki-67 as an example, stained by an antibody that fluoresces in the CY5 channel. More detailed steps involved in obtaining nuclei which are positive for ki-67 in each colony are represented in figure 2.7.

Since Ki-67 is a nuclear protein, we require DAPI-stained nuclei (figure 2.7

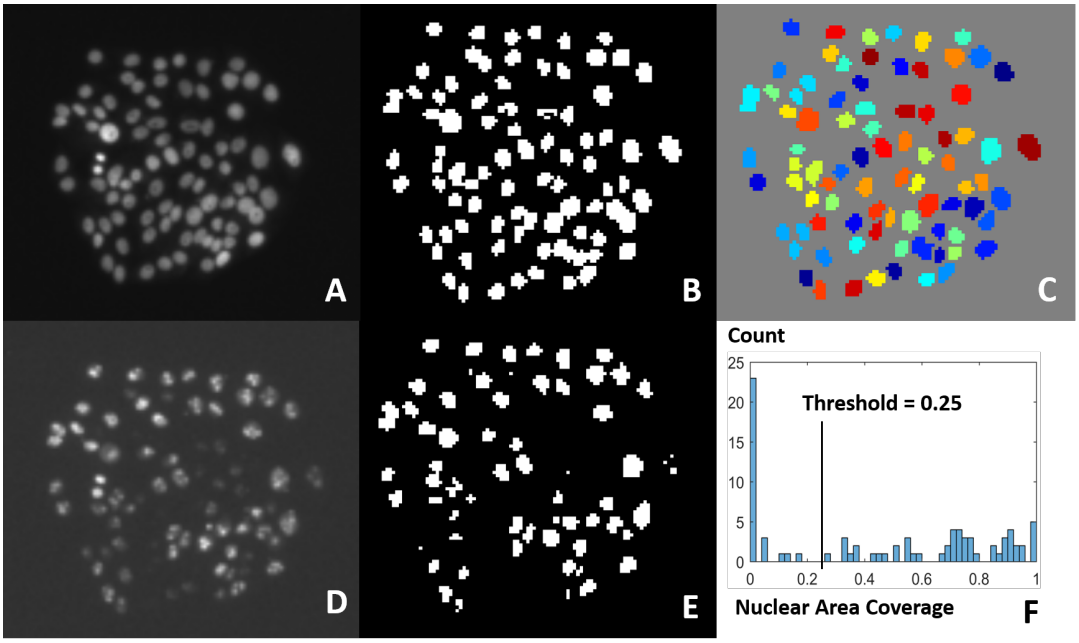

Figure 2.7: **A:** DAPI channel image of nuclear-stained colony. **B:** Binarized colony image. **C:** Segmentation of nuclei based on distance transform and watershed segmentation. **D:** CY5 channel image showing immunostained Ki-67. **E:** Binarized Ki-67 image showing Ki-67+ regions. **F:** Histogram of colony nuclei binned by fractional Ki-67 nuclear staining. All nuclei with more than 25% Ki-67 staining are marked 'Ki-67 positive'.

A). The nuclei are segmented using a distance-transform followed by watershed algorithm, similar to the scheme described previously for segmenting whole colonies (figure 2.7 B, C). The pixels constituting each nucleus are compared with those (at the corresponding location) in the binarized Ki-67 image (figure 2.7 E). Thus the fraction of total pixels (or area) positive for Ki-67 are obtained for each nucleus. If this fraction is greater than a fixed arbitrary threshold (which is 0.25 in our assay, as shown in figure 2.7 F), the nucleus is considered to be positive for Ki-67. Thus, we can obtain the fraction of cells that express ki-67 in each colony. The distribution of this metric can be obtained for the entire population, and differences between cell-lines can be analyzed.

2.4.7 Image Processing Steps in Live-cell Mode

The live-cell assay has the advantage that it requires fewer materials, no sample preparation and lesser imaging time as compared to fixed-cell mode. It is a dynamic assay, which allows colony growth to be modelled over time. Since there is no nuclear staining or immunofluorescence staining, only the fluorescent markers present in the cell-line *ab initio* are imaged - in our case, DsRed and GFP, imaged in the RFP and GFP channels.

The image processing scheme followed is similar to the fixed-cell mode in that we again employ a distance transform to identify 'basins', which are used to segment colonies by applying the watershed algorithm. The difference lies in the channels to which the segmentation step is applied. In the fixed-cell mode, images displaying nuclear staining, obtained the DAPI

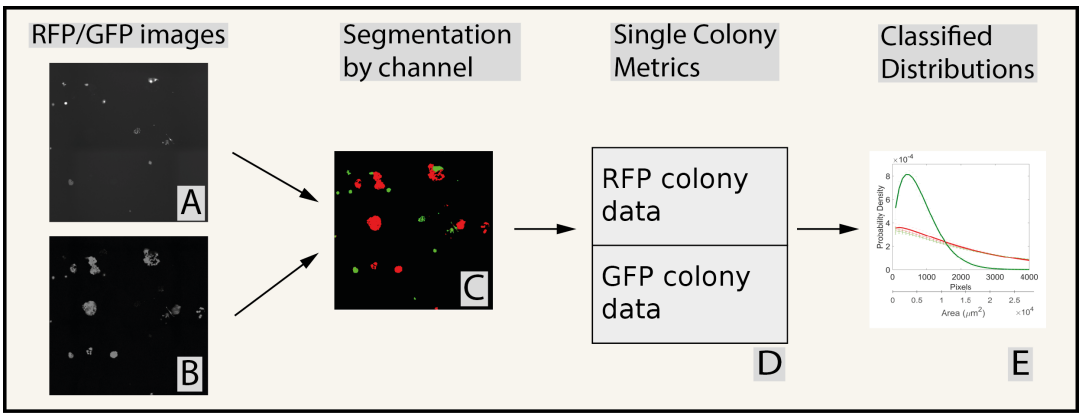

Figure 2.8: A,B: RFP/GFP channel images after preprocessing. **C:** Individual channels segmented by distance transform, watershed algorithm, and the results are combined. **D:** Colony metrics are obtained for each channel. **E:** Distributions of colony metrics are plotted together for comparison between cell lines.

channel, are used to identify individual colonies, and signal from the RFP/GFP masks enables classification of each colony as 'red' or 'green'. In the live-cell mode, we directly segment colonies using images obtained from the RFP/GFP channel (i.e. displaying the native DsRed/ GFP fluorescence of the cell-lines being used).

This effectively means that the colonies are already 'sorted' by cell line ('red' and 'green'), and the function of the algorithm here is to accurately segment individual colonies and quantify their size, area or intensities as may be required. The live-cell scheme is shown below in figure 2.8.

2.5 Animal Study

Female 5- to 7-week-old NSG mice were used according to protocols approved by the Johns Hopkins University Animal Care and Use Committee. Mice were anesthetized, and 2×10^6 MDA-MB-231 cells were injected into the mammary

fat pad (MFP) in a 1:1 ratio of DsRed- and GFP-expressing cells. In another experiment, 5×10^4 MDA-MB-231 cells were directly injected into the tail vein, in a mix of 1:1 DsRed- and GFP-expressing cells. After 3 weeks, the mice were sacrificed and the lungs were inflated with a solution of PBS:OCT (Optimal Cutting Temperature Compound, Fisher Healthcare, Pittsburgh, PA, USA), harvested and immediately fixed in 10% neutral buffered formalin for 1 hour, followed by an overnight soak in 30% sucrose solution at 4 degrees. The lungs were then frozen in OCT and cryo-sectioned into 15 um slices, that were immediately stored at -20°C.

2.6 *In Vivo* Metastatic Colony Formation

We developed an algorithm to obtain colony size distributions from fixed and stained tissue sections, based on similar principles to the *in vitro* assay described earlier.

2.6.1 Tissue Sections

Tissue slides were washed 3 times in a PBS-T solution (1% Tween in PBS 1X). In the dark, Sudan Black 0.1% was used for 25 minutes at room temperature to reduce tissue background fluorescence. After another similar washing step, the tissue slides were stained with DAPI (1:1000 in PBS) for 15 minutes at room temperature. After washing, the slides were mounted with a 90% Glycerol mounting solution and covered with a coverslip that was sealed with nail polish.

2.6.2 Imaging

Similar to the *in vitro* imaging step, slides were imaged at a magnification of 4x in the DAPI, GFP and RFP channels. A montage of images was used to cover the entire lung section on the slide, and the images produced were stitched using Gen5 (BioTek). The stitched images were reduced to 50% in size in order to facilitate processing.

2.7 Image Processing and Analysis - *In Vivo*

2.7.1 Introduction

Once the slides containing lung sections have been digitized, we can apply the image processing scripts developed in this work to obtain quantitative data from them. The full *in vivo* algorithm is given in figure 2.9. Identical treatement is performed for the RFP and GFP channels. The algorithms is divided into two parts - Part 1 segments all metastatic colonies except micrometastases, which are eliminated during the blurring step. Once the large colonies are segmented, they are masked and the remaining area is scanned for micrometastases, i.e. seeding events which are only a few cells large in size. Any such events that are within an arbitrary distance of 40 pixels from the large metastatic colonies are rejected, since we cannot know whether such a colony has been seeded from a primary tumor or represents cells migrating from the nearby metastatic colony in the organ. The DAPI channel is only used to identify lung regions from background and plays no role in the segmentation steps.

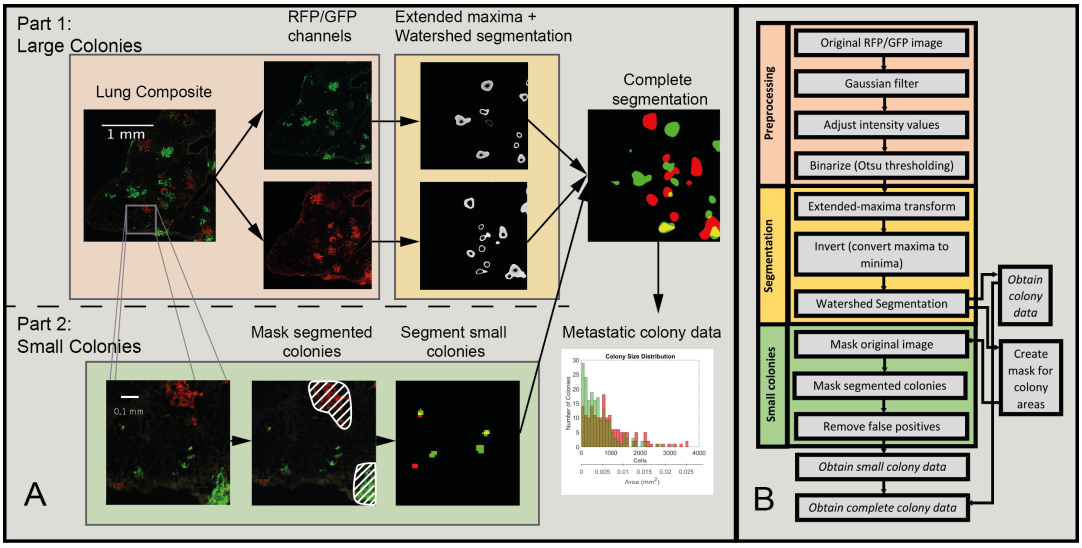

Figure 2.9: A: Overall scheme for segmenting colonies from tissue sections **B:** Flow-diagram showing steps involved

2.7.2 Preprocessing

The preprocessing steps consist of gaussian filtering and contrast adjustment. The parameters of both these filters are modified for each set of experiments to give reasonable segmentation. The gaussian filter removes micrometastases and retains the general shape of the original colonies.

2.7.3 Extended-Maximum Transform

Since the lung metastatic colonies are not circular, segmentation based on the distance transform. Hence, the extended-maximum transform is used to identify local maxima - which are regions of densely packed cells - in metastatic colonies, which serve as the foci ('basins'/ 'troughs') for the watershed algorithm. A similar method has been discussed for other applications in existing literature [46]. The segmentation is accurate for regions of dense colonies as

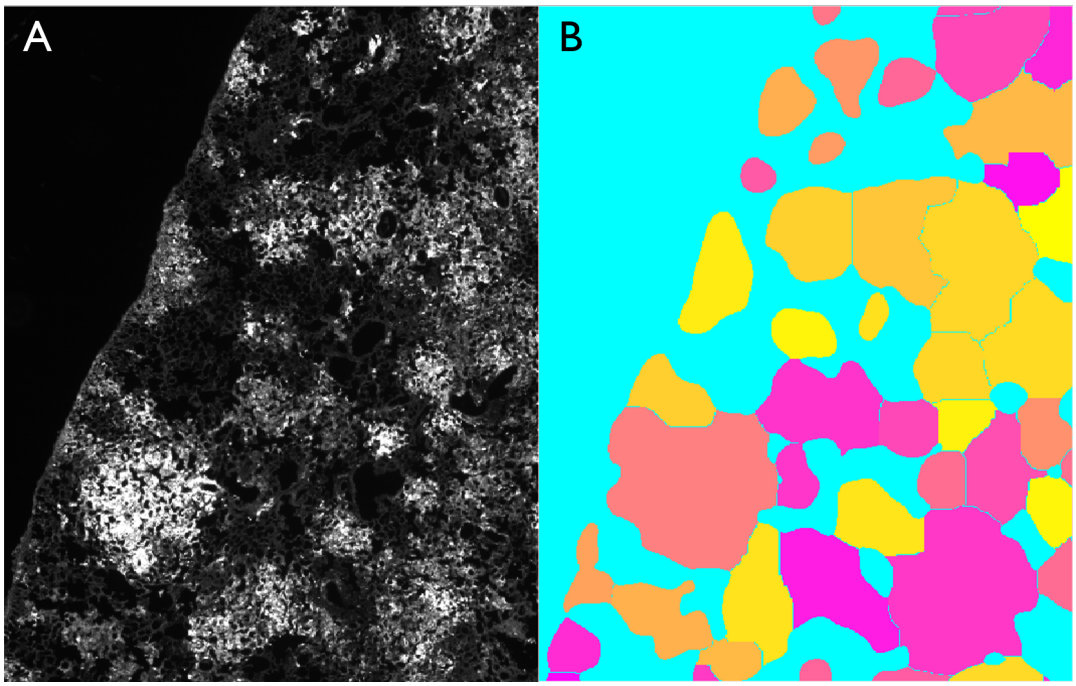

Figure 2.10: A: Region of lung with dense metastases **B:** Segmented large metastases

shown in figure 2.10. A simple threshold based segmentation step is unable to distinguish between adjacent metastases.

2.7.4 Micrometastases Segmentation

Once the larger metastatic colonies have been segmented, the intensities of all pixels that either constitute a metastatic colony (in either the RFP or GFP channel) or lie within 40 pixels of a metastatic colony, are reduced to zero. The resultant image only contains micrometastases that are reasonably far away from large metastatic colonies. Micrometastases are segmented with a simple binarization step. Artifacts due to debris or autofluorescence (due to traces of blood) were observed as bright spots positive in both GFP and RFP channels. These are eliminated by removing pixels that were positive in both GFP and

RFP channels after binarization.

2.8 Statistical Analysis

Differences in colony sizes were verified through a Wilcoxon Rank Sum test, and distributions were analyzed by the Kruskal-Wallis test. Both tests were evaluated through functions available in MATLAB or GraphPad Prism (GraphPad Software, Inc., San Diego, CA, USA).

Chapter 3

Results and Discussion

3.1 Proliferation Study of DsRed+ and GFP+ Cells

This chapter contains two sections, one each for discussing *in vivo* and *in vitro* results. Before carrying out the *in vitro* colony formation assay, we determined that there was no significant difference in the growth rates of our fluorescently labelled MCF7 cell lines by trypsinizing and counting cells in a hemocytometer over an 11 day time course, shown in Figure 3.1.

3.2 Colony Segmentation and Analysis *In Vitro*

The experimental methods and image processing algorithms used to quantify data have been discussed in the previous section. GFP+ (Green, pretreated) cells were treated with paclitaxel and co-cultured with DsRed+ cells (Red, untreated). As a control, untreated GFP+ and DsRed+ cells were co-cultured (Red, Green).

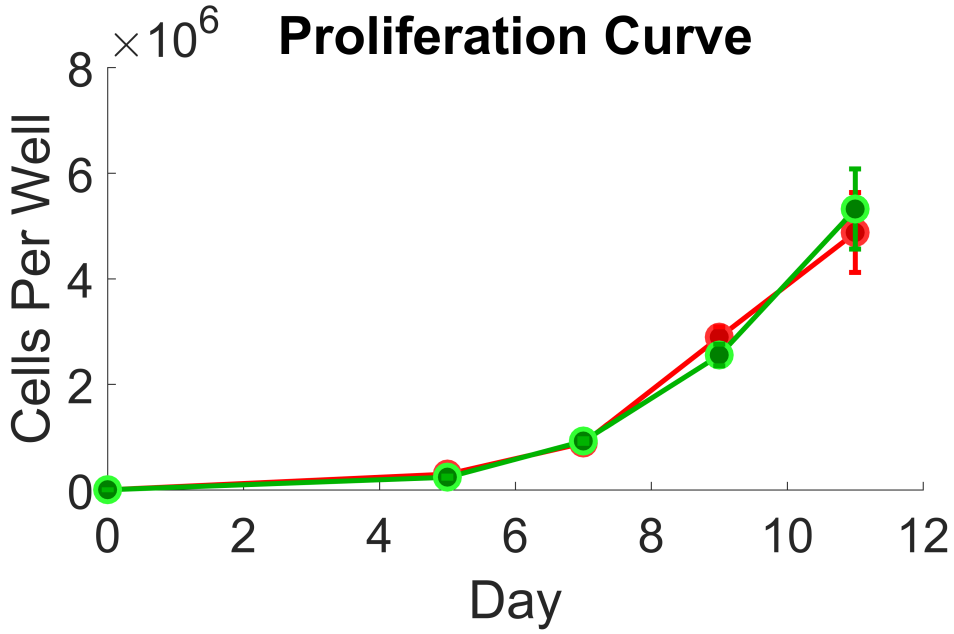

Figure 3.1: No significant differences in proliferation rates were observed

3.2.1 Live-Cell Mode: Monitoring Colony Areas

The live-cell mode is a quick and simple method that can be used to follow individual colony growth visually and monitor colony size distributions over time. The same plate can be imaged repeatedly at different time points and colony size distributions obtained. The results are shown in figure 3.2.

The total area of all the colonies can be used as a measure of confluence or proliferation of the given cell line, as shown in figure 3.2 A. Cells that had been treated with paclitaxel before plating show significantly reduced proliferation ($P<0.001$) over the experiment. The cell-line treated with paclitaxel prior to drug treatment produced smaller colonies. The colony size distribution of drug treated cells was significantly different ($P<0.001$) than those of cells without drug treatment, in co-culture or in control wells. There was no

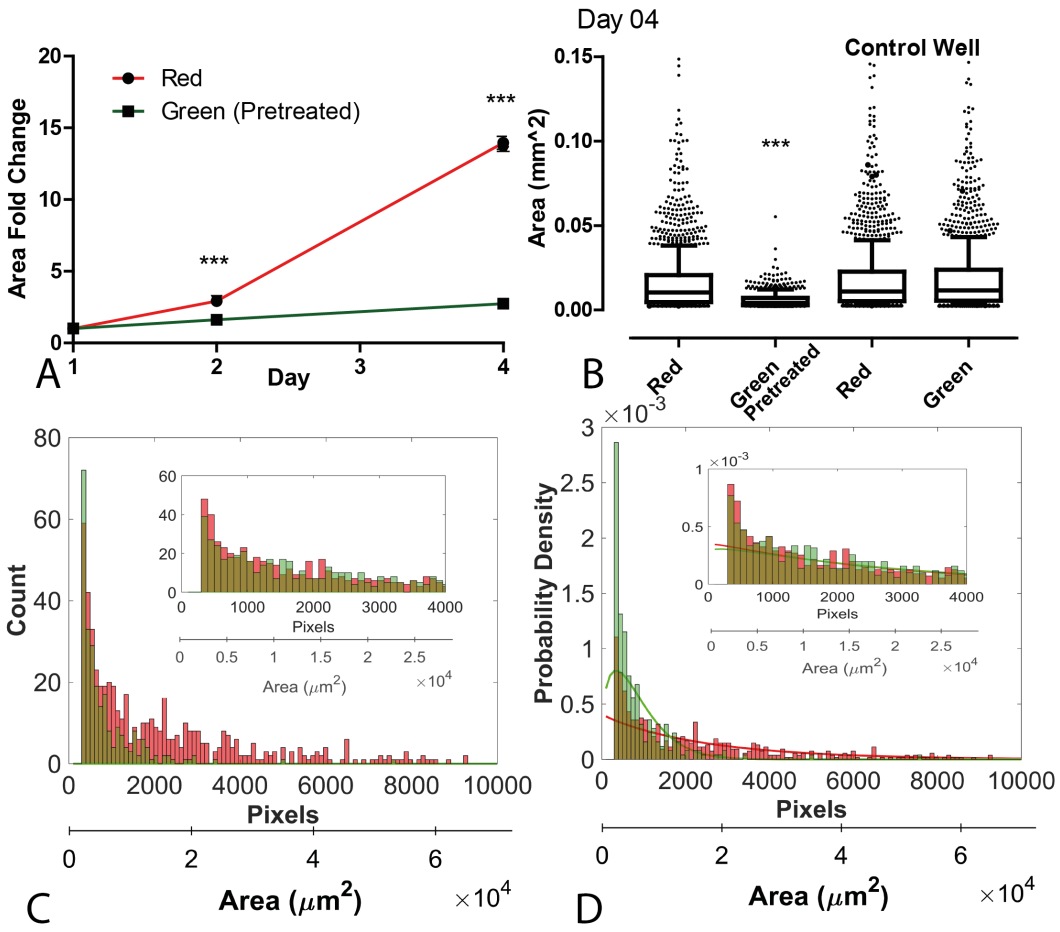

Figure 3.2: **A:** Fold change in sum of colony areas. **B:** Statistical analysis of colony size distributions on day 04, across all wells by the Kruskal-Wallis test. GFP+ cells, which were pretreated with paclitaxel prior to plating are observed to produce significantly smaller colonies, as opposed to the DsRed+ cells in co-culture. Control wells, where neither cell lines were pretreated show identical colony size distributions. **C:** Histogram showing difference in colony size distribution for DsRed+ (red) and pretreated GFP+ (green) cells, obtained from one well of a six well plate. Inset figure shows results from control well. **D:** Normalized histogram and probability density function of colony size distributions from drug-treated and (inset) control well.

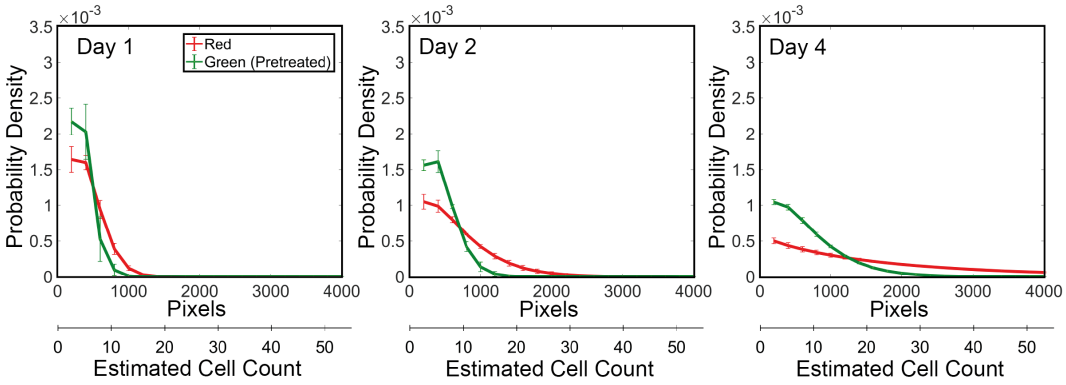

Figure 3.3: The probability density functions of colony sizes as they change over the duration of the live cell experiments. Mean and standard deviations of probability distributions functions are shown.

significant difference between the distributions of cells untreated with drug (figure 3.2B). Figures 3.2 C and D show size distributions from a single well of a six well plate, showing counts and probability densities respectively. The shift of the GFP+ histograms to the left clearly indicates smaller mean and median colony size. The DsRed+/ GFP+ histograms from control wells are indistinguishable. The probability density functions averaged over all wells, as they change over time, are showed in figure 3.3.

3.2.2 Correlating Colony Areas with Cell Counts

An important difference between the live-cell and fixed-cell modes of the *in vitro* experiments is that colony size is measured only in pixels (and not number of constituent cells) in the live-cell mode. This limitation arises from the difficulty in segmenting closely spaced cells displaying cytoplasmic staining. However, obtaining cell counts by segmenting nuclei is possible. This means that in the fixed-cell mode, which contains a nuclear-staining step,

the experimenter may obtain colony cell counts.

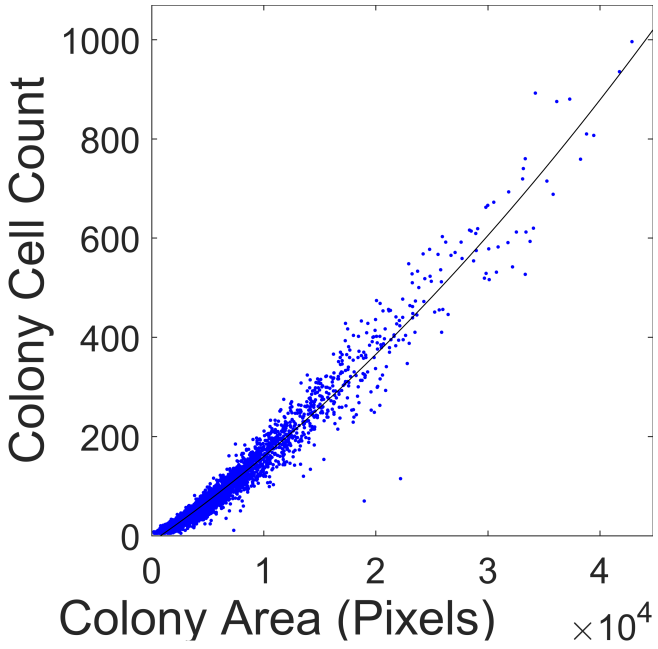

Figure 3.4: Correlating colony areas to cell counts.

Nuclei are segmented using a distance transform combined with a watershed segmentation step - similar to the algorithm used to segment individual colonies in the fixed-cell mode of the *in vitro* assay. It is then possible to obtain a correlation between colony area and cell count. Once this correlation is obtained for a given cell line, it could be applied to any live-cell experiments that may be performed, allowing the experimenter to obtain estimates of colony cell counts.

3.2.3 Fixed-cell Mode: Monitoring Colony Cell Counts and KI67 distribution

The methods involved in carrying out fixed-cell assays have been discussed previously. Similar to the effect observed in the live-cell assay, the leftward

shift in colony size distributions under effect of paclitaxel is observed in the fixed cell mode too, as shown in figure 3.5A. The image analysis scheme to investigate Ki-67 distributions have been discussed in the previous chapter. In the fixed-cell *in vitro* assay, we observed paclitaxel-treated cells to have a larger fraction of non-proliferating or low-Ki-67 colonies 3.5B.

The microscope images obtained through the fixed-cell assay on day 04 are shown in figure 3.6.

3.2.4 Validation of MATLAB tool with manual thresholding

Colony size distributions were obtained through ImageJ and compared with the results obtained by our MATLAB script. In ImageJ, binarizing RFP and GFP channel images obtained from the live-cell experiment, and analyzing all objects so obtained provided colony size distributions. While discrete colonies are segmented fairly accurately in ImageJ (or by any other simple thresholding method), connected colonies are not. Furthermore, this process needs to be repeated for every well and is time consuming as compared to our algorithm. Manual tracing of colonies would ensure accurate segmentation even for dense or connected colonies, but the amount of experimenter time taken is large (more than 20 minutes per well). Analyzing in fixed-cell mode, with nucleus counting and ki-67 evaluation is even more difficult and time consuming in ImageJ.

In figure 3.7, we compare colony size distributions obtained from the MATLAB tool described in this work and ImageJ binarization/object counting, from a single well of a 6 well-plate. While the overall shape of the distribution

remains the same, ImageJ shows elevated numbers of large colonies, probably due to inaccurate segmentation of connected colonies.

3.3 Colony Segmentation and Analysis *In Vivo*

Two experiments are discussed in this section. The first is an orthotopically transplanted tumor containing cells with the hypoxia-reporter construct. The second is a tail-vein injection of an equal mixture of DsRed+ and GFP+ cells. The details of the model and the methods involved in slide preparation and the image processing scheme to quantify metastatic colony sizes *in vivo* have been discussed in the previous chapter.

3.3.1 *In vivo* colony size distribution differences in hypoxia-exposed metastatic cells

The hypoxia-induced color changing cell lines were injected into mice mammary fat pads (MFP) to form the primary tumor. Cells exposed to intratumoral hypoxia would permanently cease expression of DsRed and begin GFP expression. Metastatic colonies formed by such a primary tumor expressed DsRed or GFP, and could be grouped into two - those originating from hypoxic cells (GFP+) and those from normoxic cells (DsRed+).

Figure 3.8 shows the results of the *in vivo* scheme of our image processing script. Metastatic colony size distributions were obtained for individual lungs (figure 3.8 C) and for the entire set used in the experiment (figure 3.8). It was observed that colonies were either DsRed+ or GFP+, indicating monoclonal origin. A significant difference was observed between the colony

size distributions produced by DsRed+ and GFP+ cells. GFP+ cells were observed to produce smaller colonies (Wilcoxon Rank-Sum test, $P<0.01$). A sample image of a lung section which did not show significant difference at the 1% level is provided, along with the probability density function in figure 3.9.

In work not yet published, it was observed that the DsRed+ cells proliferated faster than GFP+ *in vitro*, which is borne out by the greater overall metastatic load contribution by DsRed+ cells in lung sections. Furthermore, The metastatic potential of GFP+ cells vs. DsRed+ cells was compared by counting the colonies (manually, and by the algorithm presented here) and normalizing them with the number of corresponding cells present in the tumor at a given time point. It was observed that the GFP+ cells had a higher metastatic potential.

The results were validated by manual counting of the number of colonies in selected lungs and found to be consistent, as shown in figure 3.10 . Manual counting is a time consuming process and took approximately 15-25 minutes for each lung, as compared to 1-2 minutes it took for our MATLAB tool to segment the same.

3.3.2 Colony size distributions from tail-vein injection of tumor cells

Colony size distributions were obtained from experimental metastases resulting from a tail-vein injection of equal numbers of DsRed+ and GFP+ cells. Five different mice were sacrificed and lung sections from each were analyzed

to obtain colony size distributions. Furthermore, a set of four different lung sections from the same mouse were also analyzed and the results shown in figure 3.11. Compared to the orthotopic tumor injection experiment described in the previous section, the lungs had lower overall metastatic load due to the shorter time course of the experiment. No significant difference was obtained in the probability distributions between DsRed+ and GFP+ cell lines.

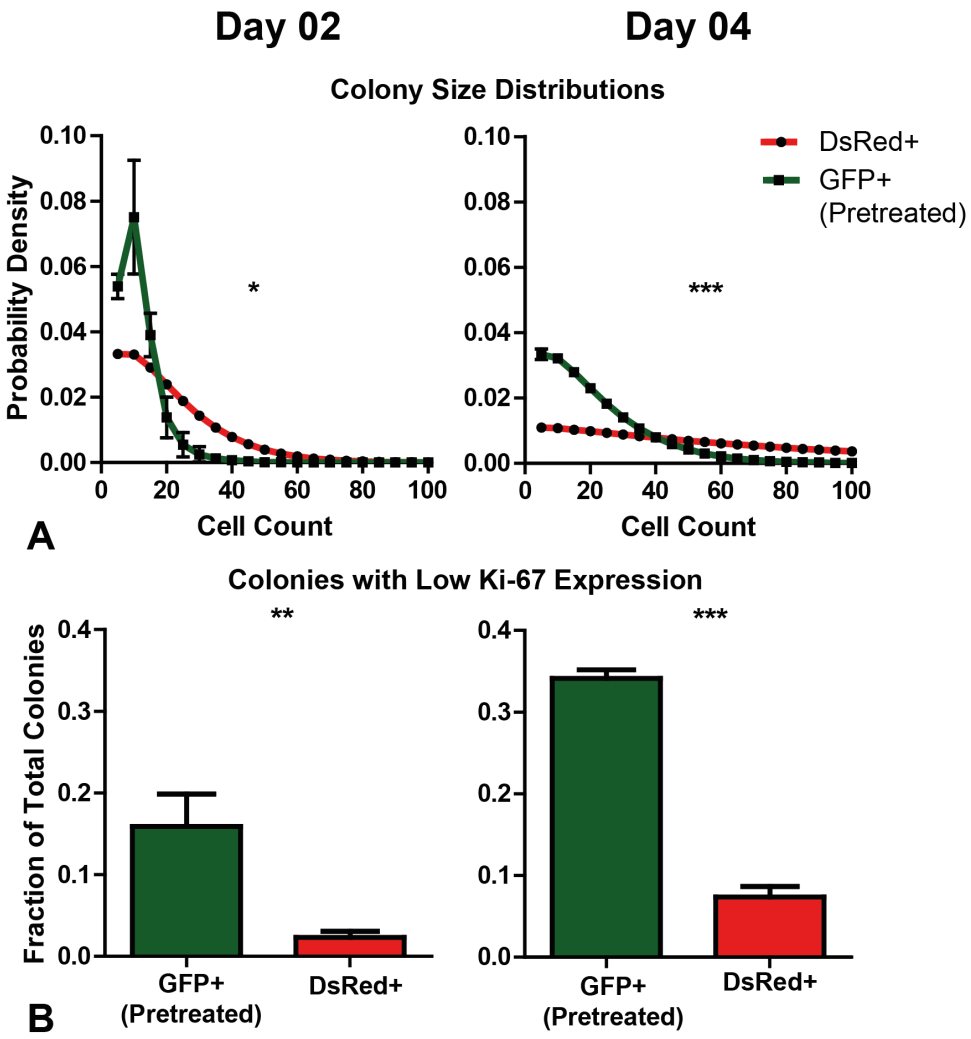

Figure 3.5: **A:** Colony size distributions over time. The Y-axis shows probability density (relative probability) and the X-axis shows colony size in number of cells. Significance was determined using the Wilcoxon rank-sum test. No significant difference was seen between DsRed+/GFP+ colony distributions in control wells (not shown) **B:** Colonies containing less than 50% of Ki-67+ cells. Treatment with Paclitaxel leads to a higher fraction of colonies having low Ki-67 expression, which is quantitatively captured by the image processing scheme. Significance was determined using 2-way ANOVA. No significant difference was seen between DsRed+/GFP+ Ki-67 expression in control wells (not shown)

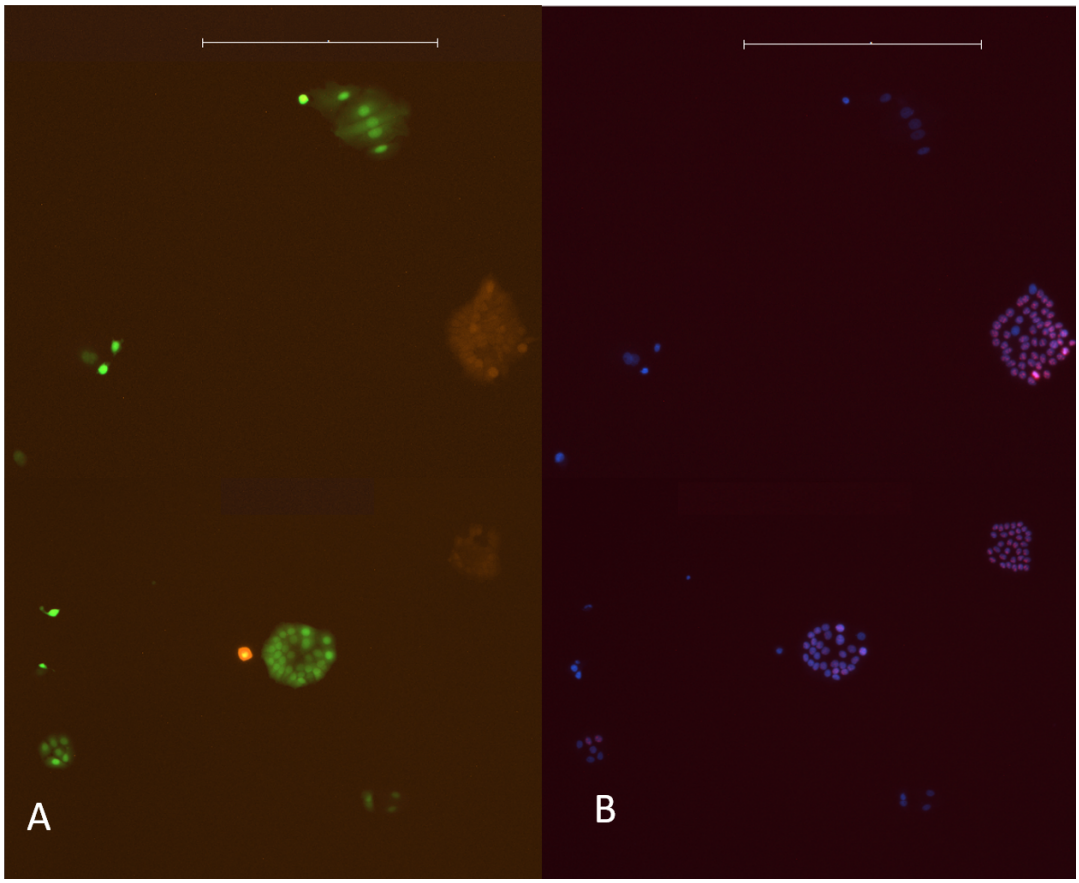

Figure 3.6: A: Microscope images showing untreated DsRed+ and pretreated GFP+ colonies. **B:** Microscope images showing DAPI (blue) and Ki-67 (red) staining. In both cases, RFP+ colonies have a greater proportion of cells displaying the Ki-67 stain. Several of the smaller GFP+ cells are devoid of Ki-67 stain entirely.

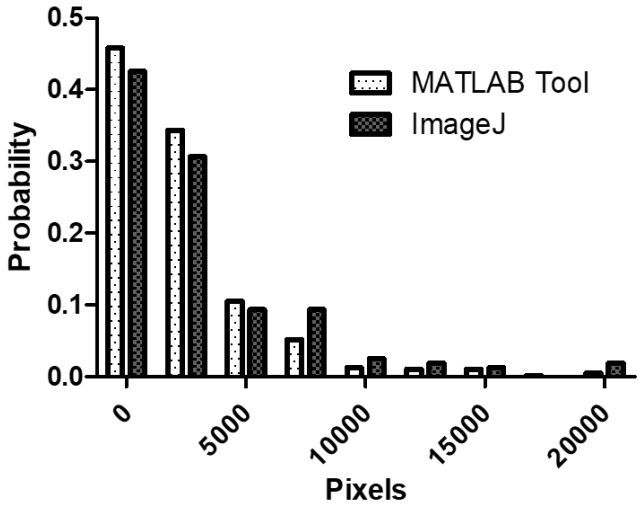

Figure 3.7: Comparison of colony size distributions obtained from the MATLAB tool described in this work and from a simple binarization/object counting step in ImageJ

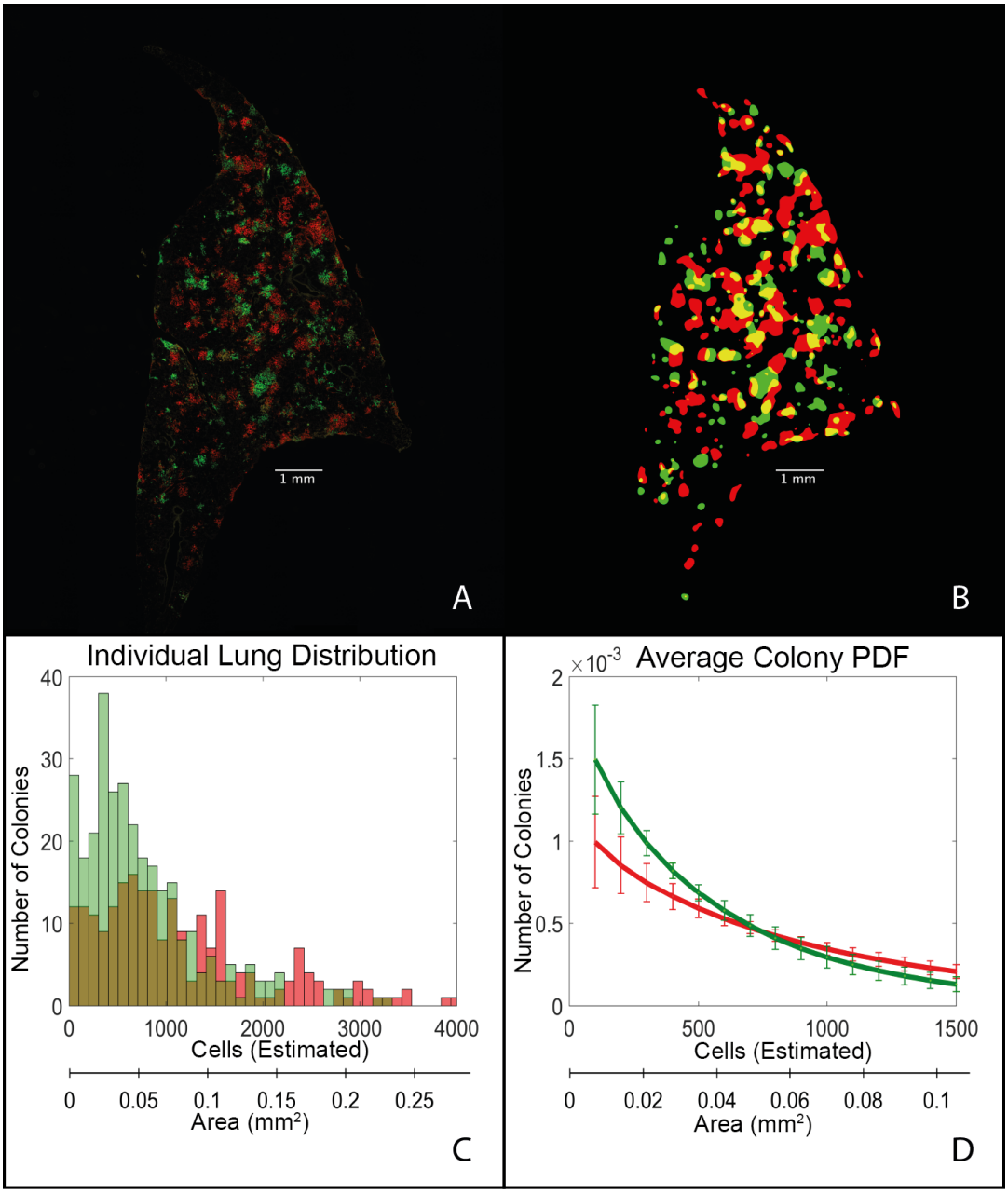

Figure 3.8: **A:** Composite lung image (RFP and GFP channels) **B:** Red and green colony regions identified. **C:** Colony size distribution obtained by analyzing a single lung section. **D:** Colony size probability distribution functions, mean and standard deviation, from our complete set of 7 lung sections.

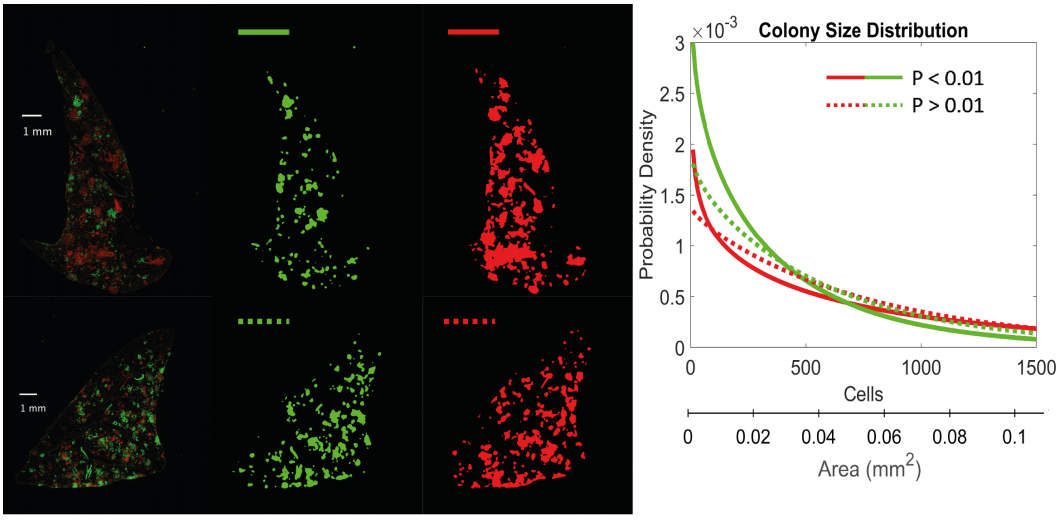

Figure 3.9: A: Representative composite (RFP+GFP) image and segmented regions of a lung that shows significant difference in colony sizes and size distribution by color. **B:** Representative images of a lung that shows no significant difference in colony sizes and size distribution.

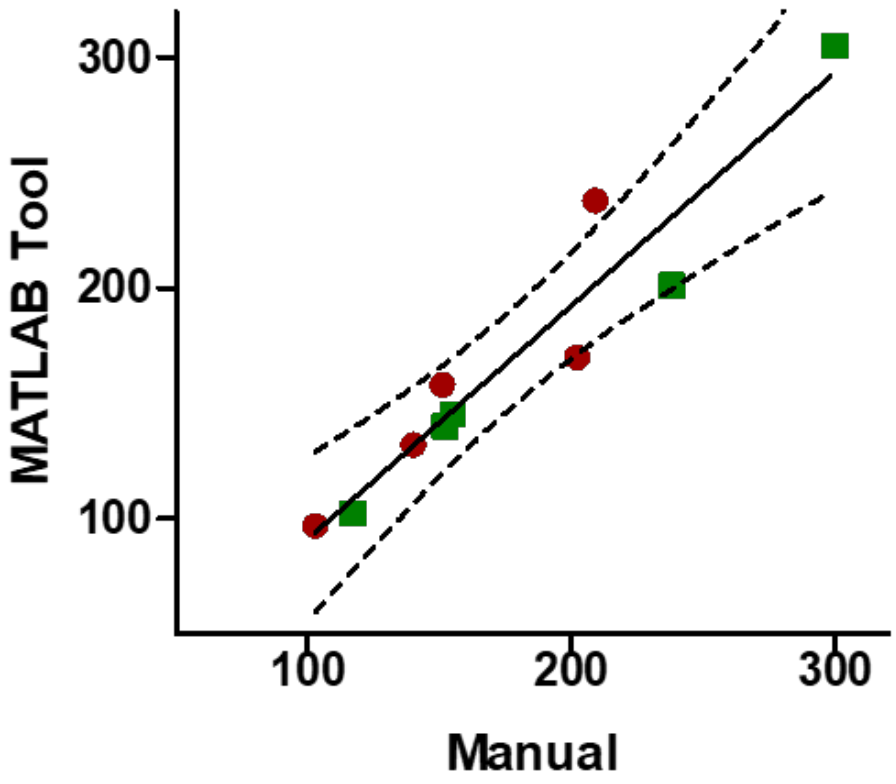

Figure 3.10: Validation of metastatic colony segmentation *in vivo*. Dashed lines show 99% confidence bounds. Red and Green colony counts are colored accordingly.

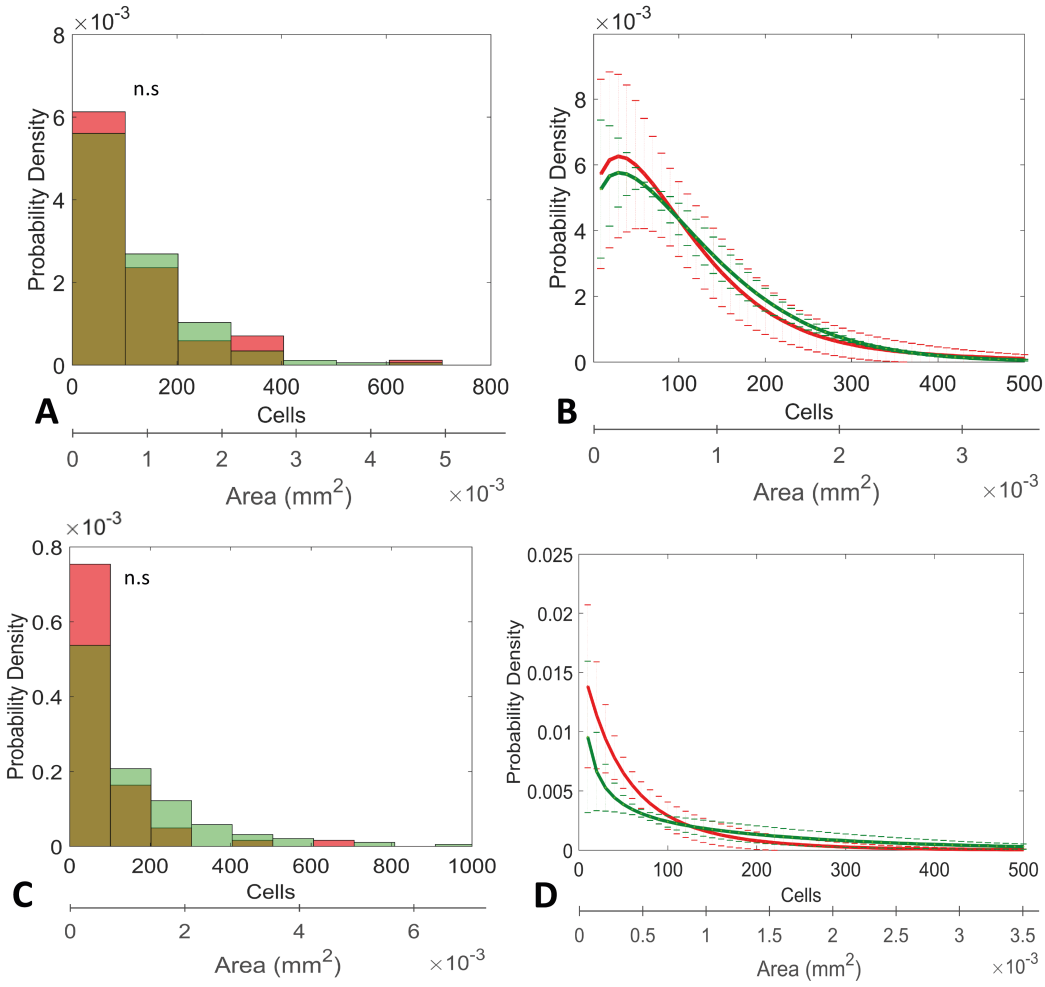

Figure 3.11: A: Consolidated probability density functions of metastatic colony sizes in lung sections of a single mouse injected with tumor cells in tail-vein. **B:** Probability density functions averaged over five lung sections of a single mouse. Error bars show standard deviation. **C:** Consolidated probability density functions of metastatic colony sizes in lung sections of five mice injected with tumor cells in tail-vein **D:** Probability density functions averaged over five lung sections of different mice. Error bars show standard deviation.

Chapter 4

Conclusion and Future Work

4.1 Conclusion

In this work we have provided an image-processing technique to obtain size distributions of colonies *in vitro* and metastatic colonies *in vivo*. We have used this method to investigate the effects of hypoxia on cancer cell metastasis, and expect it to be useful for any system of cells tagged with fluorescent markers. The method is fast and accurate even for dense colonies, and is ideal for processing a large volume of slides or high-throughput applications.

With the growing popularity of fluorescent cells and transgenic reporter mice, it is possible for the image-processing based methods described here to provide detailed and granular information on cancer clonality and metastasis both *in vitro* and *in vivo*. Such data is particularly important for developing and validating accurate mathematical models of cancer metastasis, of which there are few, one of the causes of which is the lack of substantial experimental data.

4.2 Future Work

The fixed-cell mode of the *in vitro* assay, can be modified to obtain the DNA content of each cell by quantifying the DAPI signal received from each nucleus. This would allow the cell-cycle phase (G1, G2, S) of the cell to be identified. Image analysis based cell-cycle quantification has been demonstrated in earlier work [47], and could add an interesting dimension to the colony-formation assay described in this work by allowing experimenters to monitor changes in expression of an immunostained protein-of-interest over different cell-cycle phases.

The *in vivo* assay could be expanded to include measurements of immunostained proteins-of-interest, such as Ki-67. While we did some preliminary analyses, the variability of the Ki-67 stain obtained resulted in inaccurate identification of regions of true staining as compared to background or autofluorescence in the CY5 channel.

It would be beneficial to improve segmentation accuracy and sensitivity of the algorithm, especially for different metastatic organs, such as brain, liver, lymph nodes. The liver suffers from high autofluorescence and hence segmentation accuracy is lower here. Machine-learning techniques are a promising avenue for improving segmentation, since we expect to produce a large amount of fluorescence-microscopy images which may be used as training sets. Extending this methodology to H&E stained sections would widen the scope of this work significantly. While some efforts have been taken in this direction, accurate segmentation remains challenging [48, 49].

It would be interesting to use the imaging-based assay described here (both *in vitro* and *in vivo*) to develop mathematical models of cancer metastasis. Models of cancer metastasis have used either clinical data or low-resolution metastatic data from mice, obtained using techniques such as BLI and others. The limitations of both categories have been discussed earlier. Using tumor growth, circulating-tumor-cell and colony size distribution data for our dual-colored hypoxia system over different time points, a mathematical model can be developed to characterize the experimental system. The response of such a model to changes in its parameters could shed light on the differences between hypoxic and normoxic cells. Finally, this method could be extended to characterize any such fluorescent, dual-colored system [50, 51].

Appendix A

Determining Role of TUBB6 on Proliferation and Chemoresistance

Aside from work described previously in this document, we also investigated the role of the human Tubulin Beta 6 Class V protein, coded by the TUBB6 gene, on tumor cell proliferation and chemoresistance. It was observed that TUBB6 expression was independent of Hypoxia-Inducible-Factor 1 (HIF1) and HIF2 expression, and did not significantly affect cell proliferation and chemoresistance. The TUBB6 project is discussed briefly in this appendix.

Materials and Methods

Cell Culture

HS578T cells were cultured in Dulbecco's Minimum Essential Media (DMEM) (Sigma-Aldrich, St. Louis MI - USA) supplemented with 10% Fetal Bovine Serum (FBS) and 1% penicillin-streptomycin (Invitrogen, Carlsbad) in a humidified atmosphere of 5% CO_2 at 37°C.

Knockdown by CRISPR/Cas9

We used CRISPR/Cas9 to generate TUBB6 knockdowns in HS578T cells. Lenti-CRISPR v2 plasmid used for generating a CRISPR-Cas9 endonuclease was a gift from Feng Zhang (Broad Institute, Massachusetts Institute of Technology, Cambridge, MA, obtained via Addgene (Addgene plasmid 52961)). Insert oligonucleotides that include a guide RNA sequence were designed as shown in Table A.1. After annealing, these oligos were inserted into the BsmBI cloning site. After bacterial transformation and DNA purification, all plasmid constructs were confirmed by nucleotide sequencing. The LentiCRISPR v2 plasmids and a non-targeted control were briefly co-transfected with 4 μg PsPAX2 and 1μg pMD2.G into a 10 cm dish of 293T cells using PolyJet transfection reagent (SignaGen Laboratories, Rockville, MD) according to the manufacturer's instructions. Media was refreshed 16-24h following initial transfection. Filtered viral supernatant was collected 48h post media change and was added to SUM159 and MDA-MB-231 cells. Puromycin (0.5 μg/mL) was added to the medium of cells transduced for selection. After selection, cells were expanded and used for experiments.

Western Blots

Cells were lysed in IGEPAL CA-630 buffer (150mM NaCl, 1% IGEPAL CA-630, 50 mM Tris-HCl, pH 8.0 and protease inhibitors) for 10 minutes on ice, centrifuged for 10 min at 13,000 rpm at 4°C and the insoluble debris were discarded. Whole cell lysates were fractionated by 12% SDS-PAGE and transferred to nitrocellulose membrane (Bio-Rad). The membrane was

Table A.1: CRISPR Knockdowns Used

Label	Forward Sequence
TUBB6 (site 3)	CACCGGGGTGAGCTCGGGCACGGTC
TUBB6 (site 4)	CACCGACCGTGGCCACCGTGTTCCG
TUBB6 (site 5)	CACCGGCCGCCTGCGATCCGCGCCA
TUBB6 (site 6)	CACCGTTCCTCCCCGTCATTGGCGG
TUBB6 (site 7)	CACCGCATGCCCTCGCCCAAGGTGT
HIF-1α (site 1)	CACCGCCATCAGCTATTTGCGTGTG
HIF-1α (site 2)	CACCGTGTGAGTTCGCATCTTGATA
HIF-2α (site 1)	CACCGCACCACGGTCCATACATACA
HIF-2α (site 2)	CCACGACGAATCTCCTCATGGTCGC
NTC2	CACCGCACCACGGTCCATACATACA

incubated for 1 hour with 5% milk in TBS-T (Tris-buffered saline and 0.1% Tween-20) and then incubated overnight with primary antibodies diluted in blocking buffer. Antibodies against the following proteins were used: HIF-1α (BD Biosciences, San Jose, CA), HIF-2α (Novus Biologicals, Littleton, CO), TUBB6 (ProteinTech, Rosemont, IL), TUBB3 (ProteinTech), Actin (ProteinTech), Phospho-Akt(Cell Signaling, Danvers, MA) and Akt (Cell Signaling). The membrane was then washed and incubated with the corresponding HRP-conjugated secondary antibody (Azure Biosystems, Dublin, CA) for 2 hours. After washing, the chemiluminescence signal was detected on an AZURE C300 using ECL (PerkinElmer, Waltham, MA).

Crystal Violet Proliferation Assay

To measure confluence, media was aspirated from cells in a well-plate, following which 100μm crystal violet solution was added to each well. Cells were incubated at room temperature under crystal violet solution for 10 minutes, which was sufficient time for the cells to be stained. Confluence was measured using the Cytation 5 Imaging Reader using a suitable threshold.

Results

TUBB6 knockdown in HS578T cells did not lead to significant proliferation differences

Successful knockdown of the human β tubulin V (TUBB6) was verified at the protein level by western blots shown in Figure A.1. Furthermore, knocking down HIF1 and HIF2 expression did not affect TUBB6 expression at the

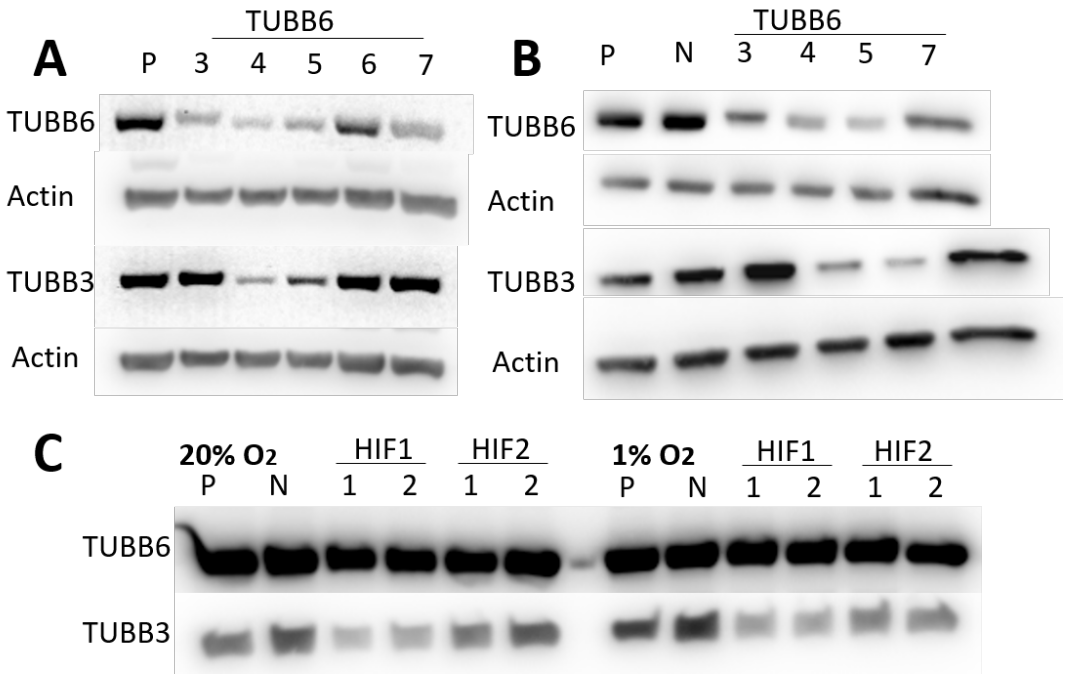

Figure A.1: A: Western blot demonstrating successful TUBB6 knockdown in CRISPR sites 3, 4, 5 and 7. Site 6 was rejected due to ineffective knockdown. Off-target TUBB3 knockdown observed in sites 4 and 5. **B:** Repeat of blot A after 30 days shows persistence of knockdown. Legend: P - Parental cell line, N - Non Target Control 2 (NTC2), 3-7 - TUBB6 CRISPR Knockdown sites, as shown in table A.1. **C:** TUBB6 expression is not governed by HIF1 or HIF2 expression or hypoxia as opposed to TUBB3. Reduced TUBB3 expression is observed in the HIF1α/ HIF2α knockdowns. Increased TUBB3 expression is also seen in parental and NTC cell lines exposed to 1%. Legend: P - Parental, N - Non Target Control 2. HIF1 and HIF2 knockdown sites are labelled.

protein level (unlike TUBB3 expression, which is in agreement with literature) . Proliferation studies were performed by measuring confluence increases using crystal violet staining. Relative changes in confluence were obtained, shown in figure A.2.

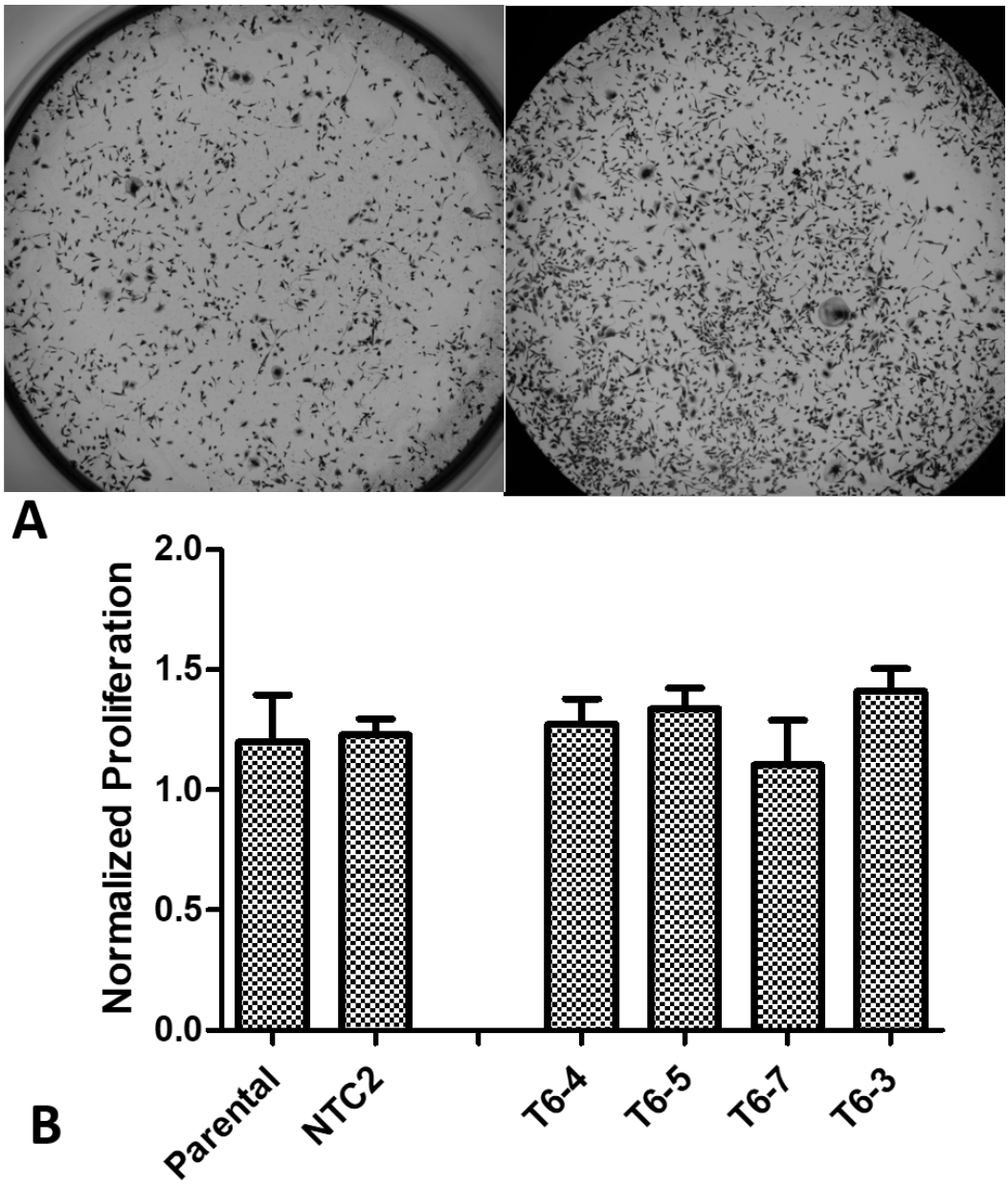

Figure A.2: **A:** Images of crystal-violet stained cells showing proliferation over a period of 48 hours. **B:** Relative increase in confluence over different knockdown cell lines, obtained by thresholding the previous images. Legend: NTC2 - Non Target Control, T6-3,4,5,6 - CRISPR knockdowns, as per table A.1

Appendix B

Supplemental MATLAB Files

In Vitro Scripts and Functions

- InVitro_Step1_ki67_Colony_Analysis.m
- InVitroo_Step2_DataProcessing.m
- InVitro_Functions_Referred.zip

In Vivo Scripts and Functions

- InVivo_Step1.m
- InVivo_Step2_AnalyzeData.m
- InVivo_Step1_TailVeinOnly.m
- InVivo_Functions_Referred.zip

Bibliography

[1] Thomas N. Seyfried and Leanne C. Huysentruyt. On the Origin of Cancer Metastasis. *Crit Rev Oncog*, 18(1-2):43–73, 2013.

[2] N. Riggi, M. Aguet, and I. Stamenkovic. Cancer Metastasis: A Reappraisal of Its Underlying Mechanisms and Their Relevance to Treatment. *Annu Rev Pathol*, October 2017.

[3] Wentao Deng, Sarah L. McLaughlin, and David J. Klinke. Quantifying spontaneous metastasis in a syngeneic mouse melanoma model using real time PCR. *Analyst*, 142(16):2945–2953, August 2017.

[4] Joan Chang and Janine T. Erler. Quantification of Lung Metastases from In Vivo Mouse Models. In *Tumor Microenvironment*, Advances in Experimental Medicine and Biology, pages 245–251. Springer, Cham, 2016. DOI: 10.1007/978-3-319-26666-4_14.

[5] Tanja Schneider, Franz Osl, Thomas Friess, Hubertus Stockinger, and Werner V. Scheuer. Quantification of human Alu sequences by real-time PCR âĂŞ an improved method to measure therapeutic efficacy of anti-metastatic drugs in human xenotransplants. *Clin Exp Metastasis*, 19(7):571–582, November 2002.

[6] Andries Zijlstra, Rebecca Mellor, Giano Panzarella, Ronald T. Aimes, John D. Hooper, Natalia D. Marchenko, and James P. Quigley. A quantitative analysis of rate-limiting steps in the metastatic cascade using human-specific real-time polymerase chain reaction. *Cancer Res.*, 62(23):7083–7092, December 2002.

[7] Melissa A. Abt, Christina L. Grek, Gautam S. Ghatnekar, and Elizabeth S. Yeh. Evaluation of Lung Metastasis in Mouse Mammary Tumor Models by Quantitative Real-time PCR. *J Vis Exp*, (107), January 2016.

[8] Maria R. Sorensen, Sara R. Pedersen, Annika Lindkvist, Jan P. Christensen, and Allan R. Thomsen. Quantification of B16 Melanoma Cells in Lungs Using Triplex Q-PCR - A New Approach to Evaluate Melanoma Cell Metastasis and Tumor Control. *PLOS ONE*, 9(1):e87831, January 2014.

[9] Ralph Weissleder and Vasilis Ntziachristos. Shedding light onto live molecular targets. *Nature Medicine*, 9(1):123, January 2003.

[10] Kazuyoshi Yanagihara, Misato Takigahira, Fumitaka Takeshita, Teruo Komatsu, Kazuto Nishio, Fumio Hasegawa, and Takahiro Ochiya. A Photon Counting Technique for Quantitatively Evaluating Progression of Peritoneal Tumor Dissemination. *Cancer Res*, 66(15):7532–7539, August 2006.

[11] Philipp M. Altrock, Lin L. Liu, and Franziska Michor. The mathematics of cancer: integrating quantitative models. *Nature Reviews Cancer*, 15(12):730, December 2015.

[12] Jacob G. Scott, Philip Gerlee, David Basanta, Alexander G. Fletcher, Philip K. Maini, and Alexander RA Anderson. Mathematical modeling of the metastatic process. *arXiv:1305.4622 [q-bio]*, May 2013. arXiv: 1305.4622.

[13] Philip Gerlee. The Model Muddle: In Search of Tumor Growth Laws. *Cancer Res*, 73(8):2407–2411, April 2013.

[14] Miguel ÃĄngel Medina. Mathematical modeling of cancer metabolism. *Critical Reviews in Oncology/Hematology*, 124:37–40, April 2018.

[15] Alex Simmons, Pamela M. Burrage, Dan V. Nicolau, Sunil R. Lakhani, and Kevin Burrage. Environmental factors in breast cancer invasion: a mathematical modelling review. *Pathology*, 49(2):172–180, February 2017.

[16] Paola Lecca and Daniele Morpurgo. Modelling non-homogeneous stochastic reaction-diffusion systems: the case study of gemcitabine-treated non-small cell lung cancer growth. *BMC Bioinformatics*, 13(Suppl 14):S14, September 2012.

[17] B. You, C. Meille, D. Barbolosi, B. Tranchand, J. Guitton, C. Rioufol, A. Iliadis, and G. Freyer. A mechanistic model predicting hematopoiesis and tumor growth to optimize docetaxel + epirubicin (ET) administration in metastatic breast cancer (MBC): Phase I trial. *JCO*, 25(18_suppl):13013–13013, June 2007.

[18] M. a. J. Chaplain, S. R. McDougall, and A. R. A. Anderson. Mathematical modeling of tumor-induced angiogenesis. *Annu Rev Biomed Eng*, 8:233–257, 2006.

[19] Niklas Hartung, Severine Mollard, Dominique Barbolosi, Assia Benabdallah, Guillemette Chapuisat, Gerard Henry, Sarah Giacometti, Athanassios Iliadis, Joseph Ciccolini, Christian Faivre, and Florence Hubert. Mathematical Modeling of Tumor Growth and Metastatic Spreading: Validation in Tumor-Bearing Mice. *Cancer Res*, 74(22):6397–6407, November 2014.

[20] K Iwata, K Kawasaki, and N Shigesada. A Dynamical Model for the Growth and Size Distribution of Multiple Metastatic Tumors. *Journal of Theoretical Biology*, 203(2):177–186, March 2000.

[21] D Barbolosi, F Verga, B You, A Benabdallah, F Hubert, Cédric Mercier, Joseph Ciccolini, and C Faivre. Modélisation du risque dâĂŹévolution métastatique chez les patients supposés avoir une maladie localisée. *Oncologie*, 13(8):528, 2011.

[22] Leonid Hanin, Jason Rose, and Marco Zaider. A stochastic model for the sizes of detectable metastases. *Journal of Theoretical Biology*, 243(3):407–417, December 2006.

[23] R. BartoszyÅĎski, L. Edler, L. Hanin, A. Kopp-Schneider, L. Pavlova, A. Tsodikov, A. Zorin, and A. Y. Yakovlev. Modeling cancer detection: tumor size as a source of information on unobservable stages of carcinogenesis. *Math Biosci*, 171(2):113–142, June 2001.

[24] Jian-Lun Xu and Philip C. Prorok. Estimating a Distribution Function of the Tumor Size at Metastasis. *Biometrics*, 54(3):859–864, 1998.

[25] Robert M. Hoffman. Strategies for In Vivo Imaging Using Fluorescent Proteins. *J. Cell. Biochem.*, 118(9):2571–2580, September 2017.

[26] Robert M. Hoffman. Use of fluorescent proteins and color-coded imaging to visualize cancer cells with different genetic properties. *Cancer Metastasis Rev*, 35(1):5–19, March 2016.

[27] Norio Yamamoto, Meng Yang, Ping Jiang, Mingxu Xu, Hiroyuki Tsuchiya, Katsuro Tomita, A. R. Moossa, and Robert M. Hoffman. Determination of Clonality of Metastasis by Cell-Specific Color-Coded Fluorescent-Protein Imaging. *Cancer Res*, 63(22):7785–7790, November 2003.

[28] Yasunori Tome, Hiroyuki Tsuchiya, Katsuhiro Hayashi, Kensuke Yamauchi, Naotoshi Sugimoto, Fuminori Kanaya, Katsuro Tomita, and Robert M Hoffman. In vivo gene transfer between interacting human osteosarcoma cell lines is associated with acquisition of enhanced metastatic potential. *Journal of cellular biochemistry*, 108(2):362–367, 2009.

[29] Atsushi Suetsugu, Yosuke Osawa, Masahito Nagaki, Hisataka Moriwaki, Shigetoyo Saji, Michael Bouvet, and Robert M. Hoffman. Simultaneous color-coded imaging to distinguish cancer âĂIJstem-likeâĂİ and non-stem cells in the same tumor. *J. Cell. Biochem.*, 111(4):1035–1041, November 2010.

[30] Takaya Abe and Toshihiko Fujimori. Reporter mouse lines for fluorescence imaging. *Development, growth & differentiation*, 55(4):390–405, 2013.

[31] Alec Vaezi, Christoph Bauer, Valeri Vasioukhin, and Elaine Fuchs. Actin Cable Dynamics and Rho/Rock Orchestrate a Polarized Cytoskeletal Architecture in the Early Steps of Assembling a Stratified Epithelium. *Developmental Cell*, 3(3):367–381, September 2002.

[32] Kevin J. Cheung, Edward Gabrielson, Zena Werb, and Andrew J. Ewald. Collective Invasion in Breast Cancer Requires a Conserved Basal Epithelial Program. *Cell*, 155(7):1639–1651, December 2013.

[33] Kari R. Fischer, Anna Durrans, Sharrell Lee, Jianting Sheng, Fuhai Li, Stephen T. C. Wong, Hyejin Choi, Tina El Rayes, Seongho Ryu, Juliane Troeger, Robert F. Schwabe, Linda T. Vahdat, Nasser K. Altorki, Vivek Mittal, and Dingcheng Gao. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature*, 527(7579):472–476, November 2015.

[34] T. T. Puck and P. I. Marcus. Action of x-rays on mammalian cells. *J. Exp. Med.*, 103(5):653–666, May 1956.

[35] Nicolaas A. P. Franken, Hans M. Rodermond, Jan Stap, Jaap Haveman, and Chris van Bree. Clonogenic assay of cells in vitro. *Nat Protoc*, 1(5):2315–2319, 2006.

[36] Herbert Braselmann, Agata Michna, Julia HeÃ§, and Kristian Unger. CFAssay: statistical analysis of the colony formation assay. *Radiation Oncology*, 10:223, November 2015.

[37] Camilo GuzmÃąn, Manish Bagga, Amanpreet Kaur, Jukka Westermarck, and Daniel Abankwa. ColonyArea: An ImageJ Plugin to Automatically Quantify Colony Formation in Clonogenic Assays. *PLOS ONE*, 9(3):e92444, March 2014.

[38] Stephen S. Raab, Dana Marie Grzybicki, Janine E. Janosky, Richard J. Zarbo, Frederick A. Meier, Chris Jensen, and Stanley J. Geyer. Clinical

impact and frequency of anatomic pathology errors in cancer diagnoses. *Cancer*, 104(10):2205–2213, November 2005.

[39] Joann G. Elmore, Gary M. Longton, Patricia A. Carney, Berta M. Geller, Tracy Onega, Anna N. A. Tosteson, Heidi D. Nelson, Margaret S. Pepe, Kimberly H. Allison, Stuart J. Schnitt, Frances P. O'Malley, and Donald L. Weaver. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA*, 313(11):1122–1132, March 2015.

[40] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H. Beck. Deep Learning for Identifying Metastatic Breast Cancer. *arXiv:1606.05718 [cs, q-bio]*, June 2016. arXiv: 1606.05718.

[41] Farzad Ghaznavi, Andrew Evans, Anant Madabhushi, and Michael Feldman. Digital Imaging in Pathology: Whole-Slide Imaging and Beyond. *Annual Review of Pathology: Mechanisms of Disease*, 8(1):331–359, 2013.

[42] Shaimaa Al-Janabi, AndrÃl' Huisman, and Paul J Van Diest. Digital pathology: current status and future perspectives. *Histopathology*, 61(1):1–9, July 2012.

[43] Liron Pantanowitz. Digital images and the future of digital pathology. *J Pathol Inform*, 1, August 2010.

[44] David C. Wilbur. Digital Cytology: Current State of the Art and Prospects for the Future. *ACY*, 55(3):227–238, 2011.

[45] Jos B.T.M. Roerdink and Arnold Meijster. The Watershed Transform: Definition, Algorithms and Parallelization Strategies. *Fundamenta Informaticae*, 41(1/2):187, January 2000.

[46] Yibo Qin, Wei Wang, Wei Liu, and Ning Yuan. Extended-Maxima Transform Watershed Segmentation Algorithm for Touching Corn Kernels. *Advances in Mechanical Engineering*, 5:268046, January 2013.

[47] Vassilis Roukos, Gianluca Pegoraro, Ty C. Voss, and Tom Misteli. Cell cycle staging of individual cells by fluorescence microscopy. *Nat. Protocols*, 10(2):334–348, February 2015.

[48] Zhenzhou Wang and Haixing Li. Generalizing cell segmentation and quantification. *BMC Bioinformatics*, 18:189, March 2017.

[49] Jacob S. Sarnecki, Kathleen H. Burns, Laura D. Wood, Kevin M. Waters, Ralph H. Hruban, Denis Wirtz, and Pei-Hsun Wu. A robust nonlinear tissue-component discrimination method for computational pathology. *Laboratory Investigation*, 96(4):450, April 2016.

[50] Gerald M. Saidel, Lance A. Liotta, and Jerome Kleinerman. System dynamics of a metastatic process from an implanted tumor. *Journal of Theoretical Biology*, 56(2):417–434, February 1976.

[51] Lance Allen Liotta, Jerome Kleinerman, and Gerald M. Saidel. Quantitative Relationships of Intravascular Tumor Cells, Tumor Vessels, and Pulmonary Metastases following Tumor Implantation. *Cancer Res*, 34(5):997–1004, May 1974.

Soumitra Bhoyar — Curriculum Vitae

Apt 1, 2923 Saint Paul Street, Baltimore, MD 21218, +1 443-889-4754, soumitra.b93@gmail.com

Education

The Johns Hopkins University, Baltimore — *Aug '16 - May '18*
MSE, Chemical and Biomolecular Eng.
GPA: 3.96 / 4.00

Institute of Chemical Technology, Mumbai — *Jul '10 - May '14*
B. Chemical Eng.
GPA: 7.88 / 10.00

Research Experience

Johns Hopkins Medical Institute, *Baltimore* — *Sep '16 - May '16*
under Dr. D. Gilkes

- Thesis research: Designed a novel MATLAB scheme to quantitatively analyze cancer metastasis in mice tissue.
- Developed a novel image-processing based assay to quantify colony growth dynamics and protein distribution in vitro.
- Determined the role of human beta-6 tubulin on proliferation, drug-responsiveness and cell-cycle of breast cancer cells.
- Experienced in cell-culture, CRISPR-Cas9, plasmid preparation in *E. Coli*, western blotting, microscopy, image processing, and statistics for data analysis.

Professional Experience

Honeywell UOP, *Delhi* — *Aug '14 - May '16*
Design Engineer

- Responsible for creating simulations and designing process equipment worth $0.5-1 million, for commercial oil refineries.
- Six-Sigma Green Belt certified under Honeywell: Improved existing work-flow to decrease man-hours required in LPG Desulfurization units.
- Supported technical discussions with customers, field specialists and project management.

United Phosphorous Ltd., *Ankleshwar* — *May '13 - Jun '13*
Engineering Intern

- Successfully identified and resolved issues faced by an underperforming belt filter platform in an agro-chemicals plant.

Additional Experience/ Projects

- **Teaching Assistant: Spring '17, Spring '18** - Teaching assistant for the courses *Project Design* and *Process Design with Aspen* at Johns Hopkins University. Leveraged professional experience to create course content and advise students in process engineering.
- **Vortex: Design Competition, 2014** - Won first prize for conceptualizing and designing a process to separate Aminoethyl piperazine from Diethylene triamine, for Dow India Ltd.
- **Azeotropy: Design Competition, 2013** - Won second prize for designing a flocculation/ adsorption unit to defluoridate contaminated groundwater, for Ion Exchange India Ltd.

Awards and Extra-curriculars

- Honeywell's Bravo award, for commendable performance on a non-standard project.
- First Prize, Prof. N. R. Kamath Chemical Engineering Quiz, by Indian Institute of Chemical Engineers, 2013.
- Experienced in high-altitude trekking and allied logistics in the Himalayas.
- Obtained highest scores in India at IGCSE/ A-level physics and chemistry, 2007 & 2009